Wanda Andreoni
Sidney Yip
*Editors*

# Handbook of Materials Modeling

## Methods: Theory and Modeling

*Second Edition*

Springer

# Handbook of Materials Modeling

Wanda Andreoni • Sidney Yip
Editors

# Handbook of Materials Modeling

Methods: Theory and Modeling

Second Edition

With 470 Figures and 32 Tables

Springer

*Editors*
Wanda Andreoni
Institute of Physics
Swiss Federal Institute of Technology – Lausanne
Lausanne, Switzerland

Sidney Yip
Department of Nuclear Science and Engineering
Department of Materials Science and Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

*We wish to dedicate this book to the memory of Walter Kohn and Aneesur Rahman, founders of the modern approach to the physics and chemistry of real materials, both theoretical and computational. While their contributions will forever stand as milestones in the progress of science, we wish to commemorate their courage and vision in proposing new methods that would be fully implemented only decades later. We – and everyone who has known them personally – continue to remember them also for their tireless enthusiasm for science, their generous sharing of ideas, their unique kindness, and their charisma. As their articles start to disappear from citation lists in the same way as those of the founders of quantum mechanics, it is both humbling and comforting to know that any present and future development rests on the work of two wonderful people who have truly embodied the values of science.*

*Wanda Andreoni and Sidney Yip*

# Preface to *HMM2*

The *Handbook of Materials Modeling*, published in 2005, was the first reference work of its kind, providing a comprehensive description of an emerging field of research and helping to grow a community at the intersection of *Computational Science* and *Materials Science and Technology* through the development of the theory, modeling, and simulation of materials across the disciplines of physics, chemistry, and engineering. In producing the first edition, the editors were guided by a simple vision – target students, young researchers, and nonspecialists as the primary audience, introduce the concept and practice of linking materials research at different resolution levels (from the atomic to the meso- to the macroscale), and stimulate exploration and applications. The e-book version became available two years later.

The *Handbook of Materials Modeling Second Edition* (*HMM2*) follows its predecessor in serving the broad community of *Computational Science* and *Materials Science and Technology*. Its goal is to update the current state of the art of materials modeling and simulation in two complementary ways:

- *Strengthen* the foundations of materials theory, modeling, and simulation
- *Broaden* the scope to include challenges and opportunities of interdisciplinary interest

In both objectives, *HMM2* strives to promote open discussion of the capabilities and limitations of current methodologies and applications. The Second Edition is composed of two companion sets: *Handbook of Materials Modeling – Methods: Theory and Modeling*, and *Handbook of Materials Modeling – Applications: Current and Emerging Materials*. These will be abbreviated as *MTM* and *ACE*, respectively. Each set is a stand-alone publication with its own Table of Contents and individual chapters that are grouped under a number of parts. Both *MTM* and *ACE* are three-volume sets. There are 11 parts in *MTM* containing, beyond the Introduction, 10 sections with 88 chapters and 16 parts in *ACE* containing, beyond the Introduction, 15 sections with 120 chapters. The chapters which constitute the primary scientific content of the handbook are organized by the section editors in each set, working in coordination with the editors in chief. The section editors also provide an introductory overview of their own section, thus facilitating access to the following chapters.

Logistical issues aside, we think it is important to emphasize the holistic aspects of materials modeling and simulation, as an evolving discipline across science and technology and the complementarity of the two sets. As the community grows in scope and relevance, learning "in depth" with the aim of mastering the relevant theoretical and computational algorithms becomes more and more important. Equally, it becomes important to be able to discern unresolved materials issues for which information from simulations could "make a difference" and to properly apply and critically assess the chosen models and methods. Thus *all readers* should recognize the beneficial overlap of the two sets. In the respective introductions we will elaborate further on this point.

An undertaking of this magnitude can only succeed as a collective effort. The *Handbook of Materials Modeling Second Edition* would not be a reality without the dedicated efforts of former Springer Editor *Maria Bellantone*, to whom we are especially indebted. She brought us together and provided us with essential logistical support. We thank *Lydia Mueller* for her early guidance. Our deepest gratitude goes to *Juby George* and *Sunaina Dadhwal* for their invaluable editorial assistance, efficient communication with authors and editors, and efficient management of the Springer platform *Meteor*. Their assistance made it possible to complete the project on an admittedly ambitious schedule. That 200 plus chapters could be commissioned, manuscripts collected, revisions, etc., be undertaken and published in less than four years is an accomplishment for which the credit properly should go to the section editors and their authors. We would like to think all participants will take pride in what we have achieved together. This is a journey that will continue. We trust the community will be diligent in upholding standards and maintaining respect and collegiality among all stakeholders.

February 2020                                                                          Wanda Andreoni
                                                                                              Sidney Yip
                                                                                                Editors

# Contents

## Volume 2

# About the Editors

**Wanda Andreoni**
Institute of Physics
Swiss Federal Institute
of Technology – Lausanne
Lausanne, Switzerland

**Wanda Andreoni** is Emeritus Professor of Physics at the Swiss Federal Institute of Technology of Lausanne (EPFL). She has been involved in Computational Materials Science since the early days of her research activity, in both academic and industrial institutions, covering also teaching, organizational, and research managing positions. She is a Fellow of the American Physical Society. She was Zernike Professor at the Groningen University (NL) (2011) and CECAM Director (2009–2012). Regarding editorial work, she was Coeditor of *Europhysics Letters* (1990–1993) and Editor of *The Chemical Physics of Fullerenes 10 (and 5) Years Later*, NATO ASI Series E: Applied Sciences, Vol. 316 (Kluwer, 1996), and of *The Physics of Fullerene-Based and Fullerene-Related Materials*, Series on the Physics and Chemistry of Materials with Low-Dimensional Structures, Vol. 23 (Kluwer, 2000).

**Sidney Yip**
Department of Nuclear Science
and Engineering
Department of Materials Science
and Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA

**Sidney Yip** immigrated to the USA from China in 1950 at age 14. After receiving degrees in Mechanical Engineering and Nuclear Engineering from the University of Michigan and spending two postdoctoral years at Cornell University, he joined the Nuclear Engineering Faculty at MIT in 1965. From early research in theoretical studies of particle and fluid transport, he became broadly involved in atomistic modeling and simulation of materials. He edited the first edition of the *Handbook of Materials Modeling* as well as *Spectroscopy in Biology and Chemistry: Neutron, X-Ray, Laser* (1974) and *Materials Interfaces: Atomic-Level Structure and Properties* (1993). Other books include the monographs, *Foundations of Neutron Transport Theory* (1967), *Neutron Molecular Spectroscopy* (1968), and *Molecular Hydrodynamics* (1980), and a text, *Nuclear Radiation Interactions* (2014). He received Guggenheim Fellowship, US Senior Scientist Prize of the Alexander von Humboldt Foundation, Distinguished Alumnus Award of the University of Michigan, and the Robert Cahn Award. A Fellow of the American Physical Society, he became Professor Emeritus in 2009.

# Section Editors



**Wanda Andreoni**  Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland

**Part: Atomistic Simulations; Plenary Topics**



**Wei Cai**  Department of Mechanical Engineering, Stanford University, Stanford, CA, USA

**Part: Crystal Plasticity: From the Atomic Scale to the Macroscale**

**Ying Chen**  Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

**Part: Stochastic, Coarse-Grained Models of Materials Mechanics**



**Somnath Ghosh**  JHU Center for Integrated Structure-Materials Modeling and Simulation (CISMMS), Air Force Center of Excellence on Integrated Materials Modeling (CEIMM), US Association of Computational Mechanics (USACM), Department of Civil Engineering, Departments of Mechanical Engineering and Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA

**Part: Crystal Plasticity: From the Atomic Scale to the Macroscale**

**Eric R. Homer**  Department of Mechanical Engineering, Brigham Young University, Provo, UT, USA

**Part: Stochastic, Coarse-Grained Models of Materials Mechanics**



**Kurt Kremer**  Max Planck Institute for Polymer Research, Mainz, Germany

**Part: Soft Matter/Polymer Simulations**

**Nicola Marzari** Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

**Part: Materials Informatics**



**Danny Perez** Theoretical Division T-1, Los Alamos National Laboratory, Los Alamos, NM, USA

**Part: Long-Timescale Atomistic Simulations: Accelerated Molecular Dynamics and Adaptive Kinetic Monte Carlo**

**Angel Rubio** Theory Department, Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

Center for Computational Quantum Physics (CCQ), The Flatiron Institute, New York, NY, USA

**Part: Electronic Structure of Materials by Ab Initio Methods**



**Stefano Sanvito** School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Center for Materials Genomics, Duke University, Durham, NC, USA

**Part: Modeling Tools for Magnetism, Magnetic Materials, and Spintronics**

**John L. Sarrao**  Los Alamos National Laboratory, Los Alamos, NM, USA

**Part: Modeling of Microstructure Evolution: Mesoscale Challenges**



**Christopher A. Schuh**  Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**Part: Stochastic, Coarse-Grained Models of Materials Mechanics**

**Marius Stan** Intelligent Materials Design, Applied Materials Division, Argonne National Laboratory, Argonne, IL, USA

**Part: Modeling of Microstructure Evolution: Mesoscale Challenges**



**Blas Pedro Uberuaga** Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA

**Part: Long-Timescale Atomistic Simulations: Accelerated Molecular Dynamics and Adaptive Kinetic Monte Carlo**

**Sidney Yip**  Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**Part: Atomistic Simulations; Plenary Topics**

# Contributors

**Claas Abert** Christian Doppler Laboratory for Advanced Magnetic Sensing and Materials, Faculty of Physics, University of Vienna, Vienna, Austria

**Luis A. Agapito** Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

**Federica Agostini** Laboratoire de Chimie Physique, University Paris-Saclay, Orsay, France

**Rabih Al Rahal Al Orabi** Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA
Solvay, Design and Development of Functional Materials Department, AXEL'ONE Collaborative Platform – Innovative Materials, Saint Fons Cedex, France

**Wanda Andreoni** Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland

**Denis Andrienko** Max Planck Institute for Polymer Research, Mainz, Germany

**M. Arul Kumar** Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA

**Vilhjálmur Ásgeirsson** Science Institute of the University of Iceland, Reykjavík, Iceland

**Alán Aspuru-Guzik** Department of Chemistry and Department of Computer Science, University of Toronto, Toronto, ON, Canada
Vector Institute, Toronto, ON, Canada

**S. Aubry** Lawrence Livermore National Laboratory, Livermore, CA, USA

**Pallab Barai** Argonne National Laboratory, Lemont, IL, USA

**Charlotte S. Becquart** UMET, Unité Matériaux et Transformations, University Lille, CNRS, INRA, ENSCL, UMR 8207, Lille, France
EM2VM, Joint laboratory Study and Modeling of the Microstructure for Ageing of Materials, Lille/Moret sur Loing, France

**Jörg Behler** Institut für Physikalische Chemie, Theoretische Chemie, Universität Göttingen, Göttingen, Germany

**Laurent Karim Béland** Department of Mechanical and Materials Engineering, Queen's University, Kingston, ON, Canada

**Tristan Bereau** Theory Group, Max Planck Institute for Polymer Research, Mainz, Germany

**Irene J. Beyerlein** Mechanical Engineering Department, Materials Department, University of California, Santa Barbara, CA, USA

**Gustav Bihlmayer** Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany

**Evgeny Blokhin** Materials Platform for Data Science, Tallinn, Estonia

Tilde Materials Informatics, Berlin, Germany

**Stefan Blügel** Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany

**Marco Buongiorno Nardelli** Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Kieron Burke** Departments of Physics and of Chemistry, University of California Irvine, Irvine, CA, USA

**Giovanni Bussi** SISSA, Trieste, Italy

**Wei Cai** Departments of Mechanical Engineering, Stanford University, Stanford, CA, USA

**Camilo E. Calderon** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Arrigo Calzolari** Center for Materials Genomics, Duke University, Durham, NC, USA

CNR-NANO Research Center S3, Modena, Italy

Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

**Laurent Capolungo** Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA

**Jesús Carrete** Institute of Materials Chemistry, TU Wien, Vienna, Austria

**Fabio Caruso** Institut für Physik and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin, Germany

**Alberto Castro** ARAID Foundation and Institute for Biocomputation and Physics of Complex Systems of the University of Zaragoza, Zaragoza, Spain

**Frank T. Cerasoli** Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

**Michele Ceriotti** Laboratory of Computational Science and Modelling, Institute of Materials, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

**James R. Chelikowsky** Center for Computational Materials, Institute for Computational Engineering and Sciences, Departments of Physics and Chemical Engineering, The University of Texas at Austin, Austin, TX, USA

**Ying Chen** Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA

**Roman V. Chepulskii** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Shreyas Cholia** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Oksana Chubykalo-Fesenko** Instituto de Ciencia de Materiales de Madrid, Spanish National Research Council – CSIC, Madrid, Spain

**Emmanuel Clouet** DEN-Service de Recherches de Métallurgie Physique, CEA, Gif-sur-Yvette, France

**Marvin L. Cohen** Department of Physics, University of California at Berkeley, Berkeley, CA, USA

Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**R. Cortes-Huerto** Max Planck Institute for Polymer Research, Mainz, Germany

**Marcio Costa** Brazilian Nanotechnology National Laboratory (LNNano), CNPEM, Campinas, Brazil

**Gábor Csányi** Engineering Laboratory, University of Cambridge, Cambridge, UK

**Basile F. E. Curchod** Department of Chemistry, Durham University, Durham, UK

**Stefano Curtarolo** Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin-Dahlem, Germany

**Pino D'Amico** CNR-NANO Research Center S3, Modena, Italy

Dipartimento di Fisica, Informatica e Matematica, Universitá di Modena and Reggio Emilia, Modena, Italy

**John Dagdelen** University of California, Berkeley, CA, USA

**Kostas Ch. Daoulas** Max Planck Institute for Polymer Research, Ackermannweg 10, Mainz, Germany

**Riccardo De Gennaro** Dipartimento di Fisica, Universit'a di Roma Tor Vergata, Roma, Italy

**Umberto De Giovannini** Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

**Mauro Del Ben** Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**L. Delle Site** Institute for Mathematics, Freie Universitat Berlin, Berlin, Germany

**Piotr de Silva** Department of Energy Conversion and Storage, Technical University of Denmark, Kgs. Lyngby, Denmark

**Christophe Domain** EDF R&D, Département MMC, Les Renardières, Moret sur Loing, France

EM2VM, Joint laboratory Study and Modeling of the Microstructure for Ageing of Materials, Lille/Moret sur Loing, France

**Claudia Draxl** Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin, Germany

**Andrea Droghetti** Nano-Bio Spectroscopy Group and European Theoretical Spectroscopy Facility (ETSF), Materials Physics Center, University of the Basque Country, San Sebastian, Spain

**S. L. Dudarev** Culham Centre for Fusion Energy, UK Atomic Energy Authority, Culham Science Centre, Abingdon, UK

**Burkhard Dünweg** Theory of Polymers, Max Planck Institute for Polymer Research, Mainz, Germany

Department of Chemical Engineering, Monash University, Clayton, VIC, Australia

**Shyam Dwaraknath** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Martin Eckstein** Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

Theory Department: Theory of Correlated Systems out of Equilibrium, Max-Planck-Institut f. Struktur u. Dynamik der Materie, Hamburg, Germany

**Anter El-Azab** School of Materials Engineering, Purdue University, West Lafayette, IN, USA

**Peter Elliott** Max Planck Institute of Microstructure Physics, Halle, Germany

**Denise C. Ford** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Marco Fornari** Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA

**Christoph Friedrich** Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany

**Giulia Galli** Department of Chemistry, Institute for Molecular Engineering, The University of Chicago, Chicago, IL, USA

**Ting Ge** Department of Mechanical Engineering and Materials Science, Duke University, Durham, USA

**Timothy C. Germann** Theoretical Division (T-1), Los Alamos National Laboratory, Los Alamos, NM, USA

**Somnath Ghosh** JHU Center for Integrated Structure-Materials Modeling and Simulation (CISMMS), Air Force Center of Excellence on Integrated Materials Modeling (CEIMM), US Association of Computational Mechanics (USACM), Department of Civil Engineering, Departments of Mechanical Engineering and Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA

**Feliciano Giustino** Department of Materials, University of Oxford, Oxford, UK

Department of Materials Science and Engineering, Cornell University, Ithaca, NY, USA

**Geena Gomez** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Rafael Gómez-Bombarelli** Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**Gerhard Gompper** Institute for Advanced Simulation and Institute for Complex Systems, Forschungszentrum Jülich, Jülich, Germany

**Priya Gopal** Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA

**Eric Gossett** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Saulius Gražulis** Department of Protein-DNA Interactions, Vilnius University Institute of Biotechnology, Vilnius, Lithuania

**Gary S. Grest** Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, USA

**Michael A. Groeber** Wright Patterson Air Force Base, Air Force Research Laboratory, Dayton, OH, USA

**Eberhard K. U. Gross** Max Planck Institute of Microstructure Physics, Halle, Germany

**Andreas Grüneis** Institute for Theoretical Physics, Vienna University of Technology, Wien, Austria

**Akash Gupta** Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

TCS Research, Tata Consultancy Services, Tata Research Development and Design Centre, Pune, India

UMI 2958, Georgia Tech-CNRS, Metz, France

**Robert M. Hanson** Department of Chemistry, St. Olaf College, Northfield, MN, USA

**Gus L. W Hart** Department of Physics and Astronomy, Brigham Young University, Provo, UT, USA

**Matti Hellström** Institut für Physikalische Chemie, Theoretische Chemie, Universität Göttingen, Göttingen, Germany

**Graeme Henkelman** Department of Chemistry and the Institute for Computational and Engineering Sciences, University of Texas at Austin, Austin, TX, USA

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX, USA

**Jan Hermann** Physics and Materials Science Research Unit, University of Luxembourg, Luxembourg, Luxembourg

**David Hicks** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Kenneth Higa** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Christian Holm** Institute for Computational Physics, University of Stuttgart, Stuttgart, Germany

**Eric R. Homer** Department of Mechanical Engineering, Brigham Young University, Provo, UT, USA

**Matthew Horton** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Bing Huang** Department of Chemistry, Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), University of Basel, Basel, Switzerland

**Patrick Huck** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Linda Hung** Toyota Research Institute, Los Altos, CA, USA

**Jürg Hutter** Institut für Chemie, Universität Zurich, Zurich, Switzerland

**Olexandr Isayev** Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC, Chapel Hill, NC, USA

Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

**Michal Jahnátek** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Anubhav Jain** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Hannes Jónsson** Faculty of Physical Sciences, University of Iceland, Reykjavík, Iceland

Department of Energy Conversion and Storage, Technical University of Denmark, Lyngby, Denmark

**Surya R. Kalidindi** Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

**Daniel Karlsson** Department of Physics, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland

**Woo Kyun Kim** Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH, USA

**Marko Knezevic** Department of Mechanical Engineering, University of New Hampshire, Durham, NH, USA

**Aleksey N. Kolmogorov** Department of Physics, Binghamton University, State University of New York, Binghamton, NY, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Kurt Kremer** Max Planck Institute for Polymer Research, Mainz, Germany

**Alessandro Laio** SISSA, Trieste, Italy

International Centre for Theoretical Physics (ICTP), Trieste, Italy

**Ricardo A. Lebensohn** Fluid Dynamics and Solid Mechanics, T-3 Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA

**Yoav Lederer** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Physics, NRCN, Beer-Sheva, Israel

**Fleur Legrain** Universitè Grenoble Alpes, Grenoble, France

CEA, LITEN, Grenoble, France

**Tony Lelièvre** CERMICS, École des Ponts ParisTech, INRIA, Champs-sur-Marne, France

**Richard LeSar** Department of Materials Science and Engineering, Iowa State University, Ames, IA, USA

**Ohad Levy** Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Department of Physics, NRCN, Beer-Sheva, Israel

**Li Li** Department of Physics and Astronomy, University of California Irvine, Irvine, CA, USA

**Lin Li** Department of Metallurgical and Materials Engineering, The University of Alabama, Tuscaloosa, AL, USA

**Laalitha Liyanage** Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

**Pui-Wai Ma** Culham Centre for Fusion Energy, UK Atomic Energy Authority, Culham Science Centre, Abingdon, UK

**E. Martinez** Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA

**Nicola Marzari** Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

**Guglielmo Mazzola** Theoretische Physik, ETH Zurich, Zurich, Switzerland

**David L. McDowell** Woodruff School of Mechanical Engineering, School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA

**Michael J. Mehl** United States Naval Academy, Annapolis, MD, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Andrius Merkys** Department of Protein-DNA Interactions, Vilnius University Institute of Biotechnology, Vilnius, Lithuania

**K. Michael Salerno** Center for Materials Physics and Technology, US Navel Research Laboratory, Washington, DC, USA

**Natalio Mingo** CEA, LITEN, Grenoble, France

**Amit Misra** Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, USA

**Joseph Montoya** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Normand Mousseau** Departement de physique and Regroupement québécois sur les matériaux de pointe, Université de Montréal, Montréal, QC, Canada

Department of Physics, University de Montréal, Montréal, Canada

**Mathias C. T. D. Müller** Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany

**Pinku Nath** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Pablo Nieves** International Research Center in Critical Raw Materials and Advanced Industrial Technologies, Universidad de Burgos, Burgos, Spain

**Chandramouli Nyshadham** Department of Physics and Astronomy, Brigham Young University, Provo, UT, USA

**Serdar Öğüt** Department of Physics, University of Illinois at Chicago, Chicago, IL, USA

**Shyue Ping Ong** University of California, Berkeley, CA, USA

**Corey Oses** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Francesco Paesani** Department of Chemistry and Biochemistry, Materials Science and Engineering, and San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA

**Michele Parrinello** Department of Chemistry and Applied Biosciences, ETH Zurich, Lugano, Switzerland

Facoltà di Informatica, Instituto di Scienze Computationali, Università della Svizzera italiana, Lugano, Switzerland

Istituto Italiano di Tecnologia, Genova, Italy

**Dvora Perahia** Department of Chemistry and Department of Physics and Astronomy, Clemson University, Clemson, SC, USA

**Danny Perez** Theoretical Division T-1, Los Alamos National Laboratory, Los Alamos, NM, USA

**Eric Perim** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Kristin Persson** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

University of California, Berkeley, CA, USA

**Brandon L. Peters** Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, USA

**Fabio Pietrucci** Muséum National d'Histoire Naturelle, UMR CNRS 7590, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, Sorbonne Université, Paris, France

**Giovanni Pizzi** Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

**Jose J. Plata** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

Departamento de Química Física, Universidad de Sevilla, Sevilla, Spain

**Giacomo Po** Mechanical and Aerospace Engineering Department, University of California, Los Angeles, CA, USA

**Evdokia Popova** Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

**R. Potestio** Max Planck Institute for Polymer Research, Mainz, Germany

**M. Praprotnik** Laboratory for Molecular Modeling, National Institute of Chemistry, Ljubljana, Slovenia

Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

**Nikolay V. Prokof'ev** Department of Physics, University of Massachusetts, Amherst, MA, USA

**Laurent Proville** DEN-Service de Recherche de Metallurgie Physique, CEA, Universite Paris-Saclay, F-91191, Gif sur Yvette, France

**Kevin Rasch** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Raffaele Resta** Consiglio Nazionale delle Ricerche (CNR), Istituto Officina dei Materiali (IOM), Trieste, Italy

Donostia International Physics Center, San Sebastián, Spain

**Tommaso Rizzo** Institute for Complex Systems, Rome Unit, National Research Council of Italy, Rome, Italy

Physics Department, Sapienza University of Rome, Rome, Italy

**David Rodney** Institut Lumière Matière, Université Lyon 1, Villeurbanne CEDEX, France

**Anthony D. Rollett** Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA, USA

**Frisco Rose** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Angel Rubio** Theory Department, Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

Center for Computational Quantum Physics (CCQ), The Flatiron Institute, New York, NY, USA

**Ivan Rungger** National Physical Laboratory, Teddington, UK

**Vladimir V. Rybkin** Institut für Chemie, Universität Zurich, Zurich, Switzerland

**Stefano Sanvito** School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Center for Materials Genomics, Duke University, Durham, NC, USA

**John L. Sarrao** Los Alamos National Laboratory, Los Alamos, NM, USA

**Matthias Scheffler** Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany

**Christopher A. Schuh** Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**Wahyu Setyawan** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Shuai Shao** Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, USA

**Sangeeta Sharma** Max Planck Institute of Microstructure Physics, Halle, Germany

**Harvey Shi** Center for Materials Genomics, Duke University, Durham, NC, USA

**R. B. Sills** Sandia National Laboratories, Livermore, CA, USA

**Ilaria Siloi** Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

**Jacopo Simoni** School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

**Jens Smiatek** Institute for Computational Physics, University of Stuttgart, Stuttgart, Germany

Helmholtz Institute Münster (HI MS), Ionics in Energy Storage, Forschungszentrum Jülich GmbH, Münster, Germany

**Sandro Sorella** International School for Advanced Studies (SISSA), Trieste, Italy

**Venkat Srinivasan** Argonne National Laboratory, Lemont, IL, USA

**Maria Stamenova** School of Physics, AMBER and CRANN Institute, Trinity College, Dublin, Ireland

**Marius Stan** Intelligent Materials Design, Applied Materials Division, Argonne National Laboratory, Argonne, IL, USA

**Alexander Stukowski** Materials Modeling Division, Institute of Materials Science, Technische Universität Darmstadt, Darmstadt, Germany

**Andrew R. Supka** Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA

**Nadine O. Symonds** Department of Chemistry, Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), University of Basel, Basel, Switzerland

**Ellad B. Tadmor** Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN, USA

**Ichiro Takeuchi** Center for Nanophysics and Advanced Materials, University of Maryland, College Park, MD, USA

Department of Materials Science and Engineering, University of Maryland, College Park, MD, USA

**Ivano Tavernelli** Zurich Research Laboratory, IBM Research GmbH, Rüschlikon, Switzerland

**Richard H. Taylor** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**Pratyush Tiwary** Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA

**Alexandre Tkatchenko** Physics and Materials Science Research Unit, University of Luxembourg, Luxembourg, Luxembourg

**Cormac Toher** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Mickaël Trochet** Departement de physique and Regroupement québécois sur les matériaux de pointe, Université de Montréal, Montréal, QC, Canada

**Alexander Tropsha** Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC, Chapel Hill, NC, USA

Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

**Igor S. Tupitsyn** Department of Physics, University of Massachusetts, Amherst, MA, USA

**Blas Pedro Uberuaga** Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA

**Demet Usanmaz** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

**Antanas Vaitkus** Department of Protein-DNA Interactions, Vilnius University Institute of Biotechnology, Vilnius, Lithuania

**Omar Valsson** Max Planck Institute for Polymer Research, Mainz, Germany

**Joost VandeVondele** Scientific Software and Libraries Unit, CSCS, ETH Zürich, Zürich, Switzerland

**Kris Van Houcke** Laboratoire de Physique Statistique, Ecole Normale Supérieure, UPMC, Université Paris Diderot, CNRS, Paris, France

**Robert van Leeuwen** Department of Physics, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland

**Troy Van Voorhis** Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA

**Carla Verdi** Department of Materials, University of Oxford, Oxford, UK

**Pierre Villars** Material Phases Data System, Vitznau, Switzerland

**O. Anatole von Lilienfeld** Department of Chemistry, Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), University of Basel, Basel, Switzerland

**Arthur F. Voter** Theoretical Division T-1, Los Alamos National Laboratory, Los Alamos, NM, USA

**Rodolphe Vuilleumier** PASTEUR, Département de chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, Paris, France

**Lucas K. Wagner** Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA

**Jian Wang** Mechanical and Materials Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA

**Haihang Wang** Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

**Shidong Wang** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA

**Jan Wilhelm** Institut für Chemie, Universität Zurich, Zurich, Switzerland

**Michael J. Willatt** Laboratory of Computational Science and Modelling, Institute of Materials, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

**Roland G. Winkler** Institute for Advanced Simulation and Institute for Complex Systems, Forschungszentrum Jülich, Jülich, Germany

**Donny Winston** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Yuanyuan Xie** California State University, Fresno, CA, USA

**Junkai Xue** Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA
Center for Materials Genomics, Duke University, Durham, NC, USA

**Kesong Yang** Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA
Center for Materials Genomics, Duke University, Durham, NC, USA

**Sidney Yip** Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**R. J. Zamora** Leadership Computing Facility, Argonne National Laboratory, Lemont, IL, USA

**Shiwei Zhang** Center for Computational Quantum Physics, Flatiron Institute, New York, NY, USA
Department of Physics, College of William and Mary, Williamsburg, VA, USA

**Caizhi Zhou** Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO, USA

**Tianyu Zhu** Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA

**Nils E. R. Zimmermann** Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Eva Zurek** Department of Chemistry, State University of New York at Buffalo, Buffalo, NY, USA

# Part I
# Introduction

# Theory and Methods for Materials Modeling: An Introduction

**1**

Wanda Andreoni and Sidney Yip

## Contents

**Abstract**

The current problems facing mankind concerning energy, health, waste, and pollution have recently begun to have a strong influence on the development of materials science and start to define its main goals. In particular, the urgent need for novel materials and for more efficient processes for their synthesis is currently driving formidable research efforts, in which modeling and computer experiments play a special role. In many scenarios, multiscale materials modeling is called for because of its ability to interrelate the descriptions of a system at various length scales – ranging from the atomic (or molecular) scale (including the description of the electronic structure) to the microscopic scale and to the

W. Andreoni (✉)
Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland
e-mail: wanda.andreoni@epfl.ch

S. Yip (✉)
Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: syip@mit.edu

mesoscopic and the macroscopic scales. This coupling is expected to enable improved predictions of microstructure evolution and thereby to lead to the development of improved materials and to improved design. However, currently, ensuring the reliability of the treatment of the system at each resolution level is still a major task for computational materials science. The *Handbook of Materials Modeling* (*HMM*) had recorded the state of the art up to 2005. The present *Handbook of Materials Modeling – Methods: Theory and Modeling* (*MTM*) presents a variety of more recent algorithms for the simulation at multiple scales and also some recent successful examples of multiscale approaches. Their explanation and critical assessment is the focus of the ten sections for which we provide here a brief survey. Moreover, we emphasize three lines of research: modeling at the mesoscale, whose critical importance has recently been recognized; multiscale simulations of complex physical and chemical processes for the diagnosis of materials behavior and as part of the synthesis protocol; and the emergence of data-driven artificial intelligence strategies.

# 1 The Attributes of Materials Modeling and Simulation

The principles of materials design are rooted in the correlation of molecular structure with physical properties. From these correlations, models are formulated that are capable of predicting microstructural evolutions. Such models allow the researcher to investigate the mechanisms underlying the critical behaviors of materials and to systematically arrive at improved design.

Recognition of this principle was evident throughout the emergence and growth of the materials modeling community as one can follow through a series of reports from the US government agencies and funding organizations, as well as reference work publications (PITAC 2005; SBES 2006; NRC 2008; WTEC 2009; MGIGC 2011; DOE-BES 2012; Konings 2012), and a commentary (Yip and Short 2013). It is worthwhile to note the perceived global societal impact of information technology in these discussions and the particular challenges of the mesoscale in building multiscale materials models. The former issue has given rise to the currently emerging field of artificial intelligence in the materials innovation domain, while the latter has given awareness to the importance of modeling and simulation at the mesoscopic level (DOE-BES 2012). This is an outcome of the "tyranny of scales" manifestation (SBES 2006).

The benefits from materials innovation can be illustrated by the current progress in nanoscience and nanotechnology, a worldwide enterprise that one may compare to computational drug design. Linking methods capable of describing a system at the atomic scale – including the quantum regime – and techniques capable of describing phenomena at the mesoscale is expected to ensure that the different phases of materials innovation and development, from design to synthesis to testing to performance and lifetime evaluations, can all be simulated, analyzed, and optimized.

In the broad arena of societal impact, one can expect materials modeling and simulation to have considerable potential for societal benefits because they sit at the crossroads of computational science and materials science and technology.

To keep the various communities connected, we declare at the outset that the terms "materials modeling and simulation" and "computational materials science" are to be regarded as synonymous unless specifically stated otherwise.

We can briefly recall three attributes of materials modeling and simulation that will endure throughout the developmental and application phases (SBES 2006). Together they constitute the essence of the "intellectual glue" joining *MTM* and *ACE*.

**Exceptional bandwidth** – The conceptual basis of materials modeling and simulation encompasses all disciplines of physical sciences and even beyond. It makes no distinction between what belongs to physics, chemistry, engineering, etc. The universality of the materials modeling enterprise becomes evident when one considers what is actually involved in formulating models for simulation. If we define modeling as the "physicalization" of a concept and simulation as the numerical implementation of this concept, then it is clear that modeling and simulation lie at the basis of all scientific disciplines that involve formulating a hypothesis and testing via simulation (or experiment) and then repeating this process through understanding and subsequent manipulation (design).

**Elimination of empiricism** – A virtue of multiscale modeling and simulation is that the results from both are conceptually and operationally quantifiable. "Quantifiability" here means one can interrogate a particular operation and determine its outcome in a controlled manner, effectively akin to an ability to investigate "cause and effect." A complex phenomenon therefore can be probed one detail at a time. This is significant because of its broad implication that empirical assumptions can be systematically replaced by physically based, quantifiable descriptions.

**Visualization of phenomena** – The numerical output of a simulation generally is in the form of data on the degrees of freedom characterizing the model. The availability of this kind of data lends itself not only to direct animation and visualization but also to data analysis and subsequent visualization of the functional properties of interest. In microscopy, for example, one can obtain structural information but usually without the energetics. Through simulation one can have both. The same is true for studies of mechanisms and reaction pathways where it would be desirable to correlate spatial and energetic information.

## 2   Complementarity of Methods and Applications and *MTM* Sections

The first edition of the *Handbook of Materials Modeling* (*HMM*) consisted of individual contributions (a total of 141 chapters) and of a special section called Plenary Perspectives containing brief commentaries (39).

In contrast, as specified in the Preface, the present edition consists of two standalone sets, namely the *Handbook of Materials Modeling – Methods: Theory*

*and Modeling* (*MTM*) and *the Handbook of Materials Modeling: Current and Emerging Applications (ACE).*

This division is not only meant to make it easier to read but also, and especially, to emphasize the interplay of methodology and "best practice" and to promote a comprehensive knowledge of materials science. It is true that, currently, the availability of several powerful and user-friendly computer codes makes it possible for researchers to undertake applications without concerning themselves with a profound and detailed understanding of the methods on which such codes are based and of the several approximations implemented therein. Such an attitude would, however, obviously be detrimental for any application because its scientific value strongly depends on the capability of the researcher to master "his/her instrument," and in particular to test and clarify the range of validity of the methods he/she is using. On the other hand, an awareness of the problems that modeling and simulation are expected to help solve, namely those that experimentalists and technologists face in diverse areas of materials science, is essential for the development of useful methods.

The theoretical and computational approaches presented in *MTM* represent to a large extent the state of the art for the study of the physics, chemistry, and mechanics of (hard and soft) materials. Note that they do not include the additional procedures that are required in any application, namely, the reduction of a specific complicated real system to a virtual model that can be treated with available computational means, the choice of the theoretical and computational approaches with assessment of the level of accuracy and limitations of the calculations, and the analysis of the results aimed at understanding and providing physical insight and connection to experiment. Why are these procedures so important? Let us not forget that what we simulate is the virtual system and what we learn from our calculations is the physics of that model. Therefore, we have to be cautious when pretending that a model represents the real system. The same is true for the simulation of the dynamics of a certain process because, for example, the duration of the virtual experiment may not allow to detect the relevant mechanisms. See also our discussion in ▶ Chap. 23, "Atomistic Simulations: An Introduction."

As seen in the Table of Contents (*TOC*), *MTM* is composed of ten sections, with each section having a number of chapters, nine on average. To give the readers who are relatively new to the community an appreciation of the scientific themes of *MTM* at an introductory level, Table 1 also shows a listing of the keywords characterizing the particular theme of each of them. These are abstracted from the titles of the sections as given in the *TOC*, intended only for a very qualitative comparison with other similar listings such as that for *ACE*. In each entry of this Table, it is implicit that the overall context is methods, theory, and models for materials modeling and simulation. The keyword abstraction is but one way of characterizing the content organization of the handbook, and clearly one can consider many other ways to categorize it. To make use of Table 1, we can look at the list for *ACE* which has 15 sections, while we need to keep in mind that the overall context for that list is

**Table 1** List of keyword descriptors of topical themes for each of the ten sections in *MTM*. For full section titles and the individual chapters in each section, see the *Table of Contents*

| Plenary topics |
| --- |
| Ab initio methods |
| Atomistic simulations |
| Long time scale simulation methods |
| Magnetism, magnetic materials, and spintronics |
| Mesoscale simulation science frontier |
| Coarse-graining methods and models |
| Soft matter/polymers simulations |
| Crystal plasticity across scales |
| Materials informatics |

**Table 2** List of keywords describing the scientific focus of the 16 symposia at the *Multiscale Materials Modeling Conference, Osaka, Japan*, Oct.28–Nov. 2, 2018. In the abbreviation, the context of multiscale modeling and simulation is implied throughout

| Theory advances |
| --- |
| Radiation effects, nuclear materials |
| Crystal plasticity: from electrons to dislocation microstructure |
| Data-driven and physics-informed materials discovery and design |
| Deformation and fracture mechanisms |
| From microstructure to properties: mechanisms, microstructure, manufacturing |
| Mechanical behavior in harsh environments |
| Mechanics of polymers, soft matter, and network materials |
| Grain boundaries: dynamics, growth, plasticity |
| Heterogeneous layered media |
| Catastrophic phenomena: fracture, earthquake |
| Crystal plasticity: structure, statistics, mechanics |
| Time- and history-dependent materials properties |
| Towards experimentally relevant time scales |
| Tribology: Multi-scale, multi-physics, multi-chemistry |
| Lubrication, wear, adhesion, friction |

current and emerging applications of materials modeling and simulation. We refer the reader to our introductory remarks to *ACE* (Chap. 1, "Applications of Materials Modeling and Simulation: An Introduction").

In Tables 2, 3, and 4, we show lists of conference symposia for three recent conferences. Table 2 gives the list of 16 symposium topics in a recent international conference on multiscale materials modeling. Table 3 shows the list of the 28 symposia at the latest (2015) quinquennial Psi-k conference. Table 4 shows the list of the nine symposia at the most recent (2019) biannual series of workshops on "Total Energy and Force Methods." The correspondence between the entries in Table 1 with Tables 2, 3, and 4 should be quite clear.

**Table 3** List of the 28 symposia at the *Psi-K Conference, San Sebastian (Spain)*, *Sept.6–10, 2015*

| |
|---|
| 30 years of Car-Parrinello |
| f-Electrons |
| Correlated Electrons |
| Theoretical Spectroscopy |
| Recent Advances in Diagrammatic Methods for the Total Energy |
| Novel Density Functionals |
| Recent Developments in Density Matrix Functional Theory |
| Density-Functional Theory for Coupled Matter-Photon Systems |
| Applications of Quantum Monte Carlo Methods |
| Upscaling Electronic Structure: Reduced-Scaling and Multi-Scale Methods |
| Spin-Orbit Coupling Effects in First-Principles Quantum Transport |
| Magnetic Excitations and Magnetization Dynamics |
| Chiral Magnetism |
| First-Principles Calculations for Multiferroics and Magnetoelectrics |
| Ab Initio Statistical Mechanics |
| Topological Insulators |
| Electrochemical Energy Storage and Conversion: Solid/Liquid Interfaces |
| Materials Design |
| Machine Learning Methods in Materials Modeling |
| Hybrid Photovoltaic Materials |
| Electron Phonon Coupling and Thermoelectricity |
| Ultrafast Charge Transfer at the Nanoscale |
| Non-Linear Optics of Materials and Nanoplasmonics |
| Novel 2D Materials and Heterostructures |
| Modeling of Defect Levels |
| Transport Properties |
| Matter Under Extreme Conditions |
| Electronic Structure Theory for Biophysics |

## 3    Brief Section Overviews

Here we briefly introduce the ten *MTM* sections and refer the reader to the overviews by the section editors for a more complete and useful account.

From the list in Table 1, we see that *MTM* begins with a "Plenary Topics" section dedicated to wide perspectives as well as to subjects that can be related to several specific sections of both *MTM* and *ACE*.

The first topical section ("Electronic Structure of Materials by Ab Initio Methods" Editor: Rubio) is dedicated to advanced quantum-mechanical methods

**Table 4** List of the sessions of the *19th Workshop on Computational Physics and Materials Science: Total Energy and Force Methods, Trieste (Italy), Jan 9–11, 2019*

| |
| --- |
| Topological Materials and Phenomena |
| Strong Correlations in Quantum Materials |
| Spins, Correlations and Entanglement and Quantum Information |
| Novel Electronic Structure Methods for Ground Excited States |
| Machine Learning/Advanced Simulation Techniques |
| Novel Electronic Structure Methods for Ground and Excited States |
| Simulating Materials and Processes across Length and Time Scales |
| Utilizing Data in Computational Materials Discovery |
| 2d Materials and Many-Electron Substrate Effects |

for the calculation of the electronic structure and related properties of extended systems. Applications of the methods discussed therein permeate this handbook, both in *MTM* as components of nonempirical simulation methods (see, e.g., the next section) and in numerous implementations and practical examples included in *ACE*.

The text proceeds with the explanation of advanced algorithms for simulations using atomistic models ("Atomistic Simulations" Editors: Andreoni and Yip). In particular, emphasis is given to simulations of the dynamics of complex physical and chemical processes in condensed phases, for which considerable progress has been made over the last 15 years. Some of the methods in both the sections just mentioned are rooted in well-established approaches that were explained in *HMM*. We refer, in particular, to GW-BSE calculations for the study of electronic excited states (Louie and Rubio 2005) and DFT-based molecular dynamics (Car et al. 2005). We have also endeavored to present very recent and promising methods that are opening new horizons for computational approaches and new opportunities for applications.

The next section ("Computational Methods for Long-Timescale Atomistic Simulations" Editors: Perez and Uberuaga) emphasizes the importance of long-scale atomistic simulations and reports on well-known – widely used – direct acceleration techniques (see also Uberuaga et al. 2005 in *HMM*) and recent extensions.

The study of magnetism requires approaches on a vast range of length and time scales. The section "Modeling Tools for Magnetism, Magnetic Materials and Spintronics" (Editor: Sanvito) is especially instructive in that it comprises theoretical and computational schemes targeted at diverse scales and models, ranging from descriptions at the atomic level to the continuum. Moreover, this section is well connected to the companion section in *ACE* (Editor: Sanvito).

Multiscale methods coupling different length scales are fundamental for the study of a large number of physical systems. Many discussions are presented in *MTM* and also in *ACE* (see, in particular, "Hierarchical Materials Modeling: Mechanical Performance" Editors: Buehler and Martin-Martinez; "Modeling the Structural Development and the Mechanics of Complex Soft Materials" Editors: Del Gado and Pellenq; "Multiscale Modeling of Diseases" Editors: Dao and Karniadakis).

The "grand challenge" pertains to the "mesoscale," and the establishment of methods for a physically sound and mathematically correct connection from the micro- to the macro-scale regimes. A full section is dedicated to the related issues ("Modeling of Microstructure Evolution: Mesoscale Challenges" Editors: Sarrao and Stan). More comments will be given below (see Sect. 4.1).

A coarse-grained representation is often extremely useful especially allowing for the transfer of information from one domain to the next within multiscale methods. Coarse-grained models and the inclusion of stochastic approaches are the focus of the next section ("Stochastic, Coarse-grained Models of Materials Mechanics" Editors: Homers, Chen, Schuh). Both basic concepts and examples of implementations and applications are discussed.

The next two sections discuss systems, properties, and processes for which advanced multiscale algorithms are called for and illustrate state-of-the-art models for simulations at different length scales.

An outstanding area is that of "soft matter" to which one full section is devoted ("Soft Matter/Polymers Simulations" Editor: Kremer). An extensive description from the atomic level to the continuum level appears to be critical in the search for structure-property and ultimately structure-process-property relations.

On the other side, namely crystalline materials, modeling plastic deformation is a key topic that has also long stimulated the development of multiscale approaches (see "Crystal Plasticity: From the Atomic Scale to the Macroscale" Editors: Cai and Ghosh).

The last section ("Materials Informatics" Editor: Marzari) deals with an area of increasing interest in the community. The explicit target here is the discovery of new materials with certain desired properties, and it is based on the data-driven "knowledge" of structure-property relations. Several chapters discuss machine-learning strategies (either established or under development) and different protocols for the creation of the reference data. We remark on an interesting application to the search for novel magnetic materials in the *ACE* part of this handbook (Chap. 17, "Machine Learning and High-Throughput Approaches to Magnetism" by Sanvito et al.). We take care to indicate also other – more recently published – machine-learning algorithms with other important applications, e.g., to predict chemical reactivity of complex organic molecules (Schwaller et al. 2018) and mechanical stability of metal-organic frameworks (Moghadam et al. 2019). This section is complemented by a chapter in the "Plenary Topics" section (as mentioned above). We shall say more below (Sect. 4.3) about artificial intelligence applied to materials science, namely about its significance and the expectations for the near future.

## 4    Highlights

In the next paragraphs, we take care to emphasize three lines of research that are bound to be central to the progress of computational materials science in the next decade and more.

## 4.1    Modeling and Simulation Challenges at the Mesoscale

A question that has been noted repeatedly in assessing the merits of materials modeling and simulation is how one can exploit the understanding of molecular mechanisms to predict the functional behavior of materials at the macroscale. In the spirit of multiscale materials modeling, one can define "mesoscale science" as the process of linking microscale models and simulations to macroscale behavior. The scientific challenges of the mesoscale have received particular interest in the community (DOE-BES 2012). They have been illustrated in attempts to explain longstanding problems in materials phenomena, such as the temperature variation of the viscosity of supercooled liquids, the temporal evolution of the shear modulus in the setting of cement paste, and the effects of loading on crack propagation in stress corrosion cracking of a glass (Yip and Short 2013). In each case, there exists a common challenge in dealing with mesoscopic time scales in molecular simulation. A section has been devoted to exploring the various issues that have been raised (Editors: Sarrao and Stan).

## 4.2    Beyond Material Properties via Simulations

As mentioned above, the search for structure-property relations is at the heart of materials science and has been guiding the development of computational tools for materials design for many years. Today's great challenge is to exploit advanced computational methods for the simulation of the physical and chemical processes that "real" materials undergo: their formation (see growth, nucleation), their transformations, and eventually their degradation. Indeed, in principle, computer experiments can provide unique information about a given material, because they can explore its behavior in conditions or at scales that are not accessible by real experiments. As pointed out by Galli in the "Plenary Topics" section, one of the most ambitious goals is to render first-principles simulations part of synthesis protocols. On the other hand, in his section overview, Kremer underlines the need to establish structure-process-property relations, which is especially critical for soft-matter/polymer systems whose observed configuration depends on their history, e.g., on the preparation method.

As demonstrated, e.g., by the accomplishments of density-functional-theory-based molecular dynamics over the last 30 years, the success will depend on the synergy of ingenious theoretical and computational algorithms aided by efficient procedures of computer science and code development.

## 4.3    Artificial Intelligence: Toward a New Paradigm in Materials Modeling

In the last few years, there has been an upsurge of interest in the materials modeling community to embrace the concepts and tools of artificial intelligence, machine

learning and deep learning. As mentioned above, this handbook contains several contributions in this area and many references. A few more comments are probably necessary to point out the significance of this new way of doing research. We remark that by now, data-based artificial intelligence strategies are considered to form the fourth paradigm of computational materials science. Born simultaneously in many different labs, this new "paradigm" is however still in its infancy and a great effort is ongoing to establish accurate transparent and shareable data. There is no doubt that this formidable endeavor will continue to offer unique opportunities in bridging the communities of computer science and computational materials science to the benefit of both. On the one hand, the development and impact in materials science will continue to be bound to the progress of the traditional approaches, theory, experiment, and computation, and the degree of interaction among them. On the other hand, the application of artificial intelligence tools is bound to create new and also unsuspected challenges, e.g., for the synthesis of the virtual materials (certainly a huge quantity) that will be produced.

# References

Car R, De Angelis F, Giannozzi P, Marzari N (2005) First-principles molecular dynamics. In: Yip S (ed) Handbook of materials modeling. Springer, pp 59–76

DOE-BES (2012) US Department of Energy Basic Energy Science report from quanta to the continuum: opportunities for mesoscale science. https://www.osti.gov/biblio/1183982-from-quanta-continuum-opportunities-mesoscale-science

Konings RJM (ed) (2012) Comprehensive nuclear materials. Elsevier, Amsterdam

Louie S, Rubio A (2005) Quasiparticle and optical properties of solids and nanostructures: the GW-BSE approach. In: Yip S (ed) Handbook of materials modeling. Springer, pp 215–240

MGIGC (2011) Materials Genome Initiative for Global Competitiveness. https://www.mgi.gov/sites/default/files/documents/materials_genome_initiative-final.pdf

Moghadam PZ, Rogge SMJ, Li A, Chow C-M, Wieme J, Moharrami N, Aragones-Anglada M, Conduit G, Gomez- Gualdron DA, Van Speybroeck V, Fairen-Jimenez D (2019) Structure-mechanical stability relations of metal-organic frameworks via machine learning. Matter 1:219

NRC (2008) National Research Council report on integrated computational materials engineering: a trans- formational discipline for improved competitiveness and national security. http://www.nap.edu/catalog/12199.html

PITAC (2005) President's Information Technology Advisory Committee report on computational science: ensuring America's competitiveness. Available via http://go.nature.com/pXa64m

SBES (2006) Report of the National Science Foundation Blue Ribbon Panel on simulation-based engineering science: revolutionizing engineering science through simulation. http://www.nsf.gov/pubs/reports/sbes_final_report.pdf

Schwaller P, Gaudin T, Lányi D, Bekas C, Laino T (2018) "Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. Chem Sci 9:6091

Uberuaga BP, Montalenti F, Germann TC, Voter AF (2005) Accelerated molecular dynamics methods. In: Yip S (ed) Handbook of materials modeling. Springer, pp 629–648

WTEC (2009) World Technology Evaluation Center-panel report on international assessment of research and development in simulation-based engineering and science. http://www.wtec.org/sbes

Yip S, Short MP (2013) Multiscale materials modelling at the mesoscale. Nat Mater 12:774

# Part II
# Plenary Topics

# Plenary Topics: An Introduction

**2**

Wanda Andreoni and Sidney Yip

## Contents

**Abstract**

The articles of this section cover themes that we expect to usefully complement the broad collection of topics presented in this part of the Handbook (MTM). When matching the contents of other MTM sections, they provide additional information and new perspectives. Also comprehensive reviews are presented of well-established theories, namely time-dependent density-functional theory and the modern theories of polarization and orbital magnetization, especially including recent formulations. Moreover, this Plenary Topic section establishes further connections (and thus reading support) to the chapters in the companion Set on Applications (ACE).

W. Andreoni (✉)
Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland
e-mail: wanda.andreoni@epfl.ch

S. Yip (✉)
Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: syip@mit.edu

# 1     Introduction

The scope of the section on "Plenary Topics" is twofold: providing topical overviews and perspectives of general interest to the readers of the Handbook and presenting the discussion of themes that were not included in the specific sections.

Some theoretical methods that have been widely applied in the last 15 years (not reported in the first issue of this Handbook) are explained extensively. For example, this is the case for time-dependent density-functional theory on which the majority of calculations of electronic excited states, electron dynamics and electron-nuclear dynamics are currently based, especially in extended systems. One of the chapters reports on general guidelines currently followed in data-based artificial intelligence applications aimed at discovering novel materials. Other chapters elucidate the role that computational approaches currently play in materials science and are expected to play in the future.

The brief survey below is meant to guide the reader through the seven chapters.

# 2     A Brief Chapter Overview

The Section opens (Cohen: ► Chap. 3, "Modeling Solids and Its Impact on Science and Technology") with an excursion over a century going through the development of modeling-based approaches to the physics of solids and its extension to nanostructured materials. We also remark that Cohen's earlier contribution to the first edition of the Handbook (Cohen 2005) remains on the recommended reading list for students and researchers coming into the community. Based on the author's address on the occasion of the 2018 Franklin Medal award, the review presented in this section leads the reader from textbook-level to some issues on which current research is focused. It shows how computational methods have emerged and progressed from empirical to "ab initio" and emphasizes the role of computational materials science as companion to experiment.

The following chapter (Galli: ► Chap. 4, "The Long and Winding Road: Predicting Materials Properties Through Theory and Computation") explains why and how first-principles calculations are presently used to explore the complex systems relevant to contemporary technology and what are the main expectations from further method developments. It is argued that exchange of information to and from experiment continues to be essential and that the envisaged role of non-empirical simulations is that of generating unique knowledge for the design of novel materials and novel synthetic routes. Another issue touched in this article is the need for accuracy, validation, and reproducibility of the results, so as to ensure the unique role of first-principles calculations in the application of data-based artificial intelligence to materials science.

For a discussion of the application of artificial intelligence, machine-learning procedures, and the use of Big Data in the realm of materials science, we refer

the reader to the Section "Materials Informatics" (Editor: Marzari). The additional contribution reported here (Draxl and Scheffler: ▶ Chap. 5, "Big Data-Driven Materials Science and Its FAIR Data Infrastructure" ) gives a detailed account of the general guiding (FAIR) principles for the "management" of the data, machine-learning models currently in use, and several recent results. Moreover, some of the problems still to be solved are clarified and important targets of future developments are mentioned.

The so-called first-principles simulations rely on a chosen description of the electronic structure of the system under investigation. For many years now, several and diverse approximate schemes of density functional theory (DFT) have been commonly used to calculate ground state properties and to drive ab initio (Car-Parrinello and Born-Oppenheimer) molecular dynamics. On the other hand, excited state properties and electron dynamics have long been a thorny problem especially for extended systems. To date, it is fair to say that time-dependent density-functional theory (TDDFT) – within the same approximations for the exchange-correlation functional as those used in ground-state DFT calculations – has become the method of choice. This fact explains why a comprehensive and detailed review is presented in this section (Agostini, Curchod, Vuilleumier, Tavernelli, and Gross: ▶ Chap. 6, "TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter"). In particular, recent extensions of TDDFT to treat non-adiabatic electron-nuclear dynamics beyond Ehrenfest (mean field) dynamics are reviewed. For a further discussion of TDDFT, we wish to direct the reader to ▶ Chap. 40, "Time-Dependent Density Functional Theory for Spin Dynamics" by Elliott et al. in this Handbook (section "Modeling Tools for Magnetism, Magnetic Materials and Spintronics" Editor: Sanvito), where the important extension to include the electron spin degree of freedom (TDSDFT) and its interaction with a magnetic field is presented in detail. Currently, TDSDFT is the only ab initio method available for the investigation of ultrafast spin dynamics and – as shown in the chapter by Elliot et al. – has successfully been applied to complex magnetic systems.

The development of methods going beyond DFT and also TDDFT and perturbative DFT approaches like GW is presently very active. Several examples are reported in the section "Electronic Structure of Materials by Ab Initio Methods" (Editor: Rubio). Here a chapter is included that describes the auxiliary-field quantum Monte Carlo (AFQMC) (Zhang: ▶ Chap. 7, "Ab Initio Electronic Structure Calculations by Auxiliary-Field Quantum Monte Carlo") for correlated many-electron systems. We recall that the core of the method relies on the utilization of the Hubbard-Stratonovich transformation to express the quantum evolution operator – of the correlated system under study – in imaginary time as an "average" (with respect to a classical Gaussian probability distribution) of free evolution-like operators. Especially for the ground state, AFQMC appears to be a clean, efficient, and relatively "simple" algorithm to handle and implement. Extension to treat excited states is also discussed. Illustrative examples refer to lattice models as well as molecules and  solids.

The last two chapters are concerned with fundamental concepts in condensed matter physics and related theories.

Under the denomination "modern theories of polarization and orbital magnetization" go rigorous definitions for a periodic system. Explicit expressions of the measurable changes in terms of a Berry phase are obtained and make direct calculations possible. The theory of polarization was well established already more than 20 years ago, and relevant applications – within DFT – started to appear soon after, leading to advances in the understanding of piezoelectricity and ferroelectricity and calculations of related properties as well as of infrared spectra of liquids and amorphous materials. Examples can be found in the ACE component of this Handbook (Picozzi: "Multiferroic and Ferroelectric Rashba Semiconductors" (Section "Applications of Materials Modeling to Magnetism, Magnetic Materials, and Spintronics" Editor: Sanvito) and Lu and Rondinelli: "Design of New Multiferroic Oxides" (Section "Oxides in Energy and Information Technologies" Editor: Franchini and Yildiz)). Extension to orbital magnetization is more recent and has allowed one to calculate, in particular, the nuclear magnetic resonance shielding tensor and the electron paramagnetic resonance $g$-tensor (from changes of the orbital magnetization) for periodic systems. ▶ Chapter 8, "Electrical Polarization and Orbital Magnetization: The Position Operator Tamed" by Resta provides an exhaustive review especially of the conceptual background and the mathematical formulation of the methods. We also wish to direct the reader to an interesting article showing the importance of the modern theory of orbital magnetization for the study of inhomogenous magnetic systems and especially of spintronics (Hanke et al. 2016).

The physics underlying the formation of a glass continues to be a severe challenge for the theory of condensed matter. The last chapter of this section (Rizzo: ▶ Chap. 9, "Critical Phenomena in Glasses") discusses the state of the art of theoretical approaches. It focuses on the mode-coupling theory, its successes and shortcomings, and incorporates recent developments aimed at improving the present characterization and understanding of the critical behavior of glasses.

## 3 Conclusions

The chapters of the section here introduced are diverse, spanning theoretical and computational methods as well as a few illustrative "best practice" examples. Each chapter provides a clear and critical view of current research and also of future challenges. In spite of the different focus of each chapter, a few important common messages can be extracted: the need for theoretical methods of wider reach, the importance of the interplay of theory, computations and experiment, and the need for more accurate algorithms as well as for their systematic validation. The latter requirements are crucial in view of the growing demand for virtual experiments and the generation of an increasing volume of data as basis for fully machine-driven approaches.

# References

Cohen ML (2005) Concepts for modeling electrons in solids: a perspective. In: Yip S (ed) *Handbook of materials modeling*. Springer, pp 13–26

Hanke JP, Freimuth F, Nandy AK, Zhang H, Blügel S, Mokrousov Y (2016) Role of Berry phase theory for describing orbital magnetism: from magnetic heterostructures to topological orbital ferromagnets. Phys Rev B 94:121114(R)

# Modeling Solids and Its Impact on Science and Technology

**3**

Marvin L. Cohen

## Contents

**Abstract**

An attempt will be made here to describe how modeling of solids in condensed matter physics and materials science has influenced science and technology for the past 100 plus years along with several examples.

## 1 Introduction

Since the focus when modeling solids and more generally condensed matter physics (CMP) is on energies, physical sizes, and time scales that are not extremely big or extremely small, but somewhere we loosely call the "middle," it can be argued that this characteristic of CMP allows it to have many links to other branches of physics

M. L. Cohen (✉)
Department of Physics, University of California at Berkeley, Berkeley, CA, USA

Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
e-mail: mlcohen@berkeley.edu

and more generally other areas of science and engineering. In addition, the domain of CMP is very broad. It has both applied and fundamental components. This paper will focus mainly on the latter with emphasis on "modeling solids" which gave rise to many intellectual and conceptual contributions to science and to applications. The description here is centered on the development of concepts and models and on the research involving the electronic structure of materials, semiconductors, superconductors, and nanoscience. It begins by briefly discussing the development of these areas over the past hundred years, followed by examples of some current achievements and discoveries.

## 2     Basic or Applied?

William Gladstone, English Finance Minister (later Prime Minister), said to Michael Faraday, "But what is the practical use of electricity?". Faraday responded, "I don't know, but someday you may tax it." When a student asks, "Should I do basic or applied science," the appropriate answer is the same as it is for the question of whether light is a particle or a wave. The correct answer is "Yes."

We have so many questions of this type in science and mathematics. In the latter case, computer-based proofs, using the products of applied computer science, are more readily accepted in recent years. If we set aside the philosophies connected with motivation or goals in science and look at what really takes place, there are countless examples of how developments in basic science have led to marvelous technical advances and countless examples of how the discovery of new insights into nature was arrived at through the pursuit of applications. Many of the great successes in science and technology resulted from conceptual advances, sometimes on a philosophical level, ab initio developments from fundamental theories, empirical models, modeling, and simulation. There were arguments made by great scientists, like Dirac, suggesting that the beauty and simplicity of an equation are often connected with its correctness. At the other end, there are complex approaches, particularly in the biosciences, that work well but have many widgets and appendices. Even though they don't have the "elegance" of the Dirac equation, they can be profound, highly appropriate for solving the problems at hand and the genesis of new insights. Early America's most famous scientist, Benjamin Franklin, is an excellent example of someone who benefitted from the cross-fertilization between basic and applied studies. Franklin's famous experiments were designed to prove that lightning is electricity, and these experiments led to his invention of the lightning rod. In the course of his studies, he was the first to label different electrical components as positive and negative, and he was the first to discover the principle of conservation of charge. Hence the notion of there being only an "Einstein or Edison" approach to science is not well founded whether one is looking for discoveries of new principles or new technologies.

When describing the modeling of solids and its associated fields like condensed matter physics (CMP) and materials science, it is useful to point out that like Franklin, CMP embodies similar attributes. It rivals chemistry in its useful applications, and its contributions to basic physics are often at the forefront. CMP is the largest branch of physics with one foot in fundamental physics and the other in applied physics where inventions such as the transistor have had a large effect on developed and some underdeveloped societies. More than 28 Nobel Prizes have been awarded in this field, covering achievements in basic and applied areas. In addition, CMP has strong connections to most areas of science and engineering because of its focus on the sizes of energies, distances, and lengths of time that are sort of "in the middle" of those studied in physics. Since these sizes are of interest in many other fields, the result is that CMP is diverse and it is in a "sweet spot" or "Goldilocks area" of science.

## 3    Empirical and Ab Initio

In covering this time span, it is important to realize that around 1900 the existence of atoms was still debated, and research such as that represented by Einstein's thesis on sugar molecules and his work on Brownian motion were important in establishing belief in the existence of atoms. Einstein's interest in problems in CMP was certainly based on an atomic model of solids. A classic example is his paper on the heat capacity of solids. The theory of electrons produced by Paul Drude is similar in spirit, and his simple approach to transport is still widely used today. Hence the concept of a solid being composed of interacting atoms was being established at that time, and the introduction of X-ray crystallography produced the picture of a crystalline solid as a periodic array of interacting atoms. When one adds the chemical identity of the constituent atoms, this view or model of a solid allows some schemes for classifications of solids.

We will return to the conceptional models later; let us first consider the question of how to calculate properties once a conceptional model has been developed. Often the first approach can be characterized as empirical, semiempirical, or heuristic. Even Einstein's Nobel Prize winning theory of the photoelectric effect was characterized as heuristic. The Bohr model of an atom was tremendously useful for explaining experimental data, for understanding many aspects of quantum theory, and it is still considered an excellent teaching aid in atomic and quantum physics and a reliable model for making estimates of physical properties in a large number of cases. Conceptually and practically, the Bohr model has been replaced by the quantum mechanical approach where, for example, wavefunctions can be calculated and used to calculate the probabilities associated with determining the position of an electron in contrast to the purely planetary model of Bohr with electrons viewed as particles. This is a good example of an ab initio approach, such as the use of

the Schrodinger equation replacing the empirical Bohr model. A major point to be made is that the use of the empirical model in explaining and understanding the physics of atoms had a large effect on the development of the ab initio approach. Another point is that at a fairly fundamental level, the Bohr model solved a host of problems and in some sense "skimmed the cream" in the early days of applying quantum ideas to atomic physics. The photoelectric effect is another case where the heuristic theory was followed by more fundamental approaches. However, it is important to emphasize that not only did the empirical theories provide conceptual models to use when developing ab initio theories, but the empirical theories were first and therefore have the characteristics of a "breakthrough."

## 4     Models for Modeling

As was proposed above, modeling a crystalline solid as a periodic array of interacting atoms is a natural approach, and this model has become the most popular. In the end, properties are calculated based on this model, and the results are compared to experiment. Usually the comparison is made using a response function such as the reflectivity which is a representative response function for an electromagnetic probe.

Since the goal of many calculations of the properties of solids is the construction of a desired response function, another way of thinking about describing or modeling solids, often referred to as the elementary excitation model, is based on considering the response function first. Fundamentally, measuring response functions is the central task of experimentalists, and in a sense the principal job for theorists is to explain and/or predict the response functions. The concept of a response function is associated with the basic question about how we learn about the physical world. In most cases, we probe and respond. Our senses are seeing, hearing, touching, smelling, and tasting. The probes for these senses are light, vibrations, changing the temperature, using a stream of liquid or gas. By using these probes and our "built-in" response sensors, we learn about the physical world around us. Similarly, for measuring properties of solids, a few examples of the probes and their associated response functions are light (dielectric function), magnetic field (magnetic susceptibility), and temperature change (heat capacity). In the elementary excitation model, we describe a solid in terms of its excitations which we create when we probe the solid as illustrated in Fig. 1. The focus in this approach is on the responses and hence the response function.

We have described two basic models. The interacting atom model focuses on the building blocks and their interactions, while the elementary excitation model focuses on the responses to probes and the emergent excitations characterizing the solid. Both methods are important, and their applications have had a tremendous influence on technology, and the related concepts have influenced other scientific areas and science philosophy. The usefulness of the two approaches depends on what properties are being studied. Early goals were classifications of solids in terms of properties such as hardness or composition or structure. Although these

**Fig. 1** Illustration of the elementary excitation model. A probe (hammer) creates elementary excitations (phonons) measured by a response sensor (ear)
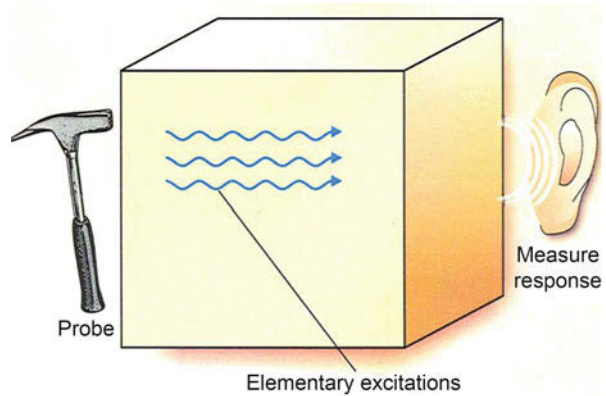


mechanical, chemical, or structural properties are important for classifying and modeling solids, a more influential property involves the electronic nature of a solid, the electrical resistivity. This is discussed next using the interacting atom model. The questions related to the building blocks in this model would be the atoms. Models of atoms such as the Bohr model or the Thomas-Fermi model can give insight, but using a full quantum mechanical approach with some approximations for potentials is the standard approach. Sometimes the building blocks are not really atoms, and the solid is considered to be just a box of free electrons. The discussion of the resistivity below illustrates the use of the model and some of the approximations.

## 5    Electronic, Optical, and Structural Properties

The range of values of resistivity is enormous. For example, consider copper with a value of $10^{-8}$ ohm-m compared to quartz with $10^{17}$ ohm-m. To emphasize this span, it has been stated that the range is the same as the comparison between the size of our galaxy and the head of a pin. Materials are classified according to their resistivities: metals with values around $10^{-8}$ ohm-m, semimetals $10^{-5}$, semiconductors $10^{-4}$ to $10^{11}$, and insulators $10^{16}$ to $10^{24}$. The resolution of the problem of finding the origin of this wide range of properties and the four classes of materials is based in the concepts related to electronic band theory of crystalline solids. The central concept is an energy gap or forbidden energy region which can arise when atoms interact. In the interacting atom model, the overlap between the atomic wavefunctions spreads or widens the energies of the sharp atomic levels into bands of allowed energy bands with energy regions or gaps where electron states are not available for occupancy. As a result one can have several scenarios. Suppose electron states are filled within each band of energies up to the energy gap. Because of the Pauli principle, the states can have at most two electrons, each in a different spin state, e.g., one with spin up and one with spin down. Because the next unoccupied state is above the gap in energy, if the energy to "jump over" this

gap is not supplied by electromagnetic radiation or heat, the electrons are frozen and won't contribute to electrical conductivity when an electric field is applied. At finite temperatures, some electrons will be free to conduct, and the number will depend on the size of the energy gap. This is a description of a semiconductor or insulator. Roughly speaking, if the gap is 3 eV or less, the material is considered to be a semiconductor. If the gap is larger, it is classified as an insulator. When bands are partially filled, the separation between an occupied and empty state is extremely small for a bulk solid, and the amount of energy required to free an electron so that it can contribute to the transport of current is minute. Hence partially filled bands lead to metallic behavior. Semimetallic behavior is the result of having bands with low densities of electrons arising from the overlapping of the bands.

To calculate the resistivity which is the response function for an electric field, we need the band structure represented by the function $E_n(k)$ where n is the band index and k represents the electron wavevector. In a sense the grouping of electrons into bands arises from the periodicity of the crystal structure. One can argue that since the cells in a crystal repeat, the effect on the electronic states is similar to the effect of confinement in a box which leads to discrete electronic states. Now the discrete electron states group into bands because of their interactions with the ionic cores. Once we have determined $E_n(k)$, we can then compute response functions resulting from electronic excitations.

How do we obtain $E_n(k)$? What models give this important function? A model introduced in the 1930s and still useful today is called the free electron gas (FEG) model where we assume that the outer electrons have been freed from the atoms and the positive cores have been smoothed out into a constant positive background "jelly." This model is sometimes called the jellium model. Because the electrons are now free to move through the solid, the model is appropriate for a metal where any excitation will allow an electron to contribute to the transport of current in the sample. If scattering is introduced empirically to this model, we have the Drude approach used to explain resistivity from a classical point of view. However now we have a quantum model, and this model can explain many quantum effects such as the electronic heat capacity, magnetism, and other properties that were mysterious in the 1930s. Starting with the second row of the periodic table, Na, Mg, and Al are metals, so this FEG model is appropriate for describing their electron properties. However, the next element is Si which is a semiconductor with bonds where valence electrons pile up between the cores as shown in Fig. 2, and a gap in $E_n(k)$ exists which is related to the energy required to liberate an electron from a bond.

As described above, semiconductors have filled bands, valence bands, empty bands, and conduction bands. The valence bands contain the electrons responsible for creating the bonds holding the crystal together. When an electron is excited to states above the gap and makes a transition into the conduction band, it is free to conduct electricity and respond to electric fields much like a metal does. The bonds and gaps arise primarily from the interaction of the valence electrons with the attractive positive cores. In the FEG, the cores were smeared out into a structureless jelly, but now the model describing a semiconductor or insulator is a periodic lattice of positive cores as shown in Fig. 2. If the potential describing the valence

**Fig. 2** Schematic picture of a semiconductor like silicon showing ionic cores (made of the nucleus and core electrons) and valence electrons forming bonds between the cores. The simple cubic structure is used here for illustration. Experimentally, silicon exists in the diamond structure at atmospheric pressures at room temperature
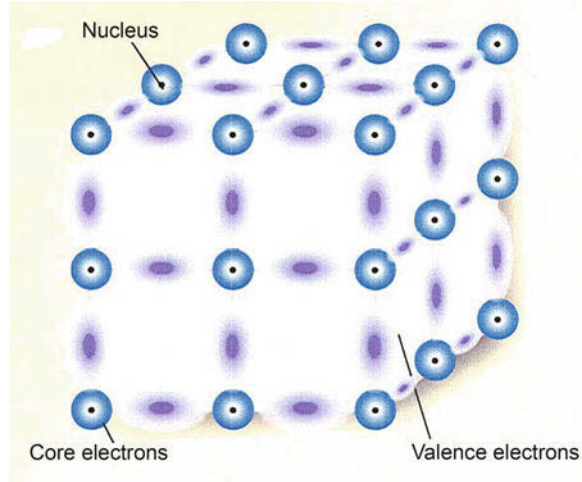
electron-core interaction is weak, the FEG won't be changed much, but the symmetry of the lattice and the nature of the potential can affect the electrons and create the bonds and gaps. At large distances from the core, a valence electron responds to the attractive Coulomb 1/r potential, but near the core, the Pauli principle is effective in keeping the valence electrons from occupying core states and keeps the valence electrons out of the regions near the core and into the bonding sites. This effect can be expressed by a repulsive potential which cancels part of the Coulomb potential and results in a net weak potential called the pseudopotential (Fermi 1934; Phillips and Kleinman 1959) which is illustrated schematically in Fig. 3. If we assume pseudopotentials at each lattice site, or more specifically, at each atomic site associated with a lattice point, then these atomic potentials perturb the FEG and produce bonds such as the covalent bonds in Si shown in Fig. 2.

The above model is the standard model for a solid. Because of the periodicity, it is convenient to express the potential as a Fourier sum in reciprocal or wavevector space. There is a one-to-one relationship between the real space lattice and a lattice in reciprocal space characterized by lattice vectors G. When the real space crystal pseudopotential is expressed as a sum of atomic pseudopotentials in a Fourier sum, the form factors for the potential are V(G). A structure factor is used to put the potentials at the proper atomic sites. Using Si as an example, the V(G) for the three smallest Gs allowed by the symmetry of the lattice, $V(G^2) = V(3), V(8)$, V(11), are large. The values for V(G) decrease for larger G, roughly representing the potential at smaller distances where the potential is weak. So all that is needed is three numbers to represent the potential. Once these are known, the FEG can be perturbed using a standard quantum mechanical approach.

The empirical pseudopotential method (EPM) (Cohen and Bergstresser 1966) is an approach in which the V(G) values are obtained from experiment. For the Si example, optical data alone provides the three needed coefficients. In fact, they are overdetermined. A similar approach is appropriate for diamond, Ge, and gray

**Fig. 3** Schematic plot of an atomic pseudopotential as a function of distance from the nucleus. The core region containing the nucleus and the core electrons is shaded. The large distance behavior of the pseudopotential is Coulombic, but this attractive potential is effectively canceled in the core region because of the Pauli exclusion principle acting on the core-valence electron interactions
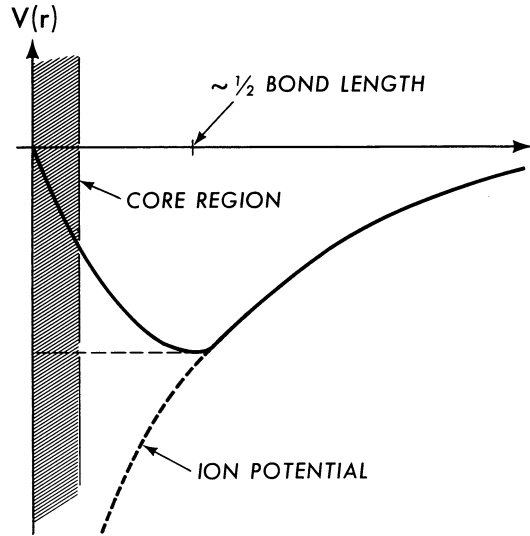
Sn since they all have the same crystal structure. Once the V(G) form factors are determined, a calculation of $E_n(k)$ is straightforward along with a calculation of the corresponding wavefunction for each state. From these one can calculate the desired response function. For example, the optical constants such as the reflectivity, transmission, or absorption describing responses to light with energies from the IR to the UV arising from electronic transitions can be obtained from a calculation of the complex frequency-dependent dielectric function using $E_n(k)$ and the wavefunctions. The results of these calculations have been impressive. Starting in the 1960s, dozens of materials were studied, and for the most part, the data were explained and successful predictions were and are common for the EPM. The method was extended to allow the valence electron density to adjust to nonperiodic changes to study surfaces and interfaces. In addition to response functions, it was possible to explore the electron density by squaring the wavefunction (Walter and Cohen 1971). The prediction of the spacially dependent electron density was later verified using X-ray scattering data.

In the above work, the crystal structure of the solid is taken from experiment. To calculate the crystal structure from first principles, an ab initio pseudopotential (ABP) is generated starting with atomic wavefunctions. There are many methods for creating the ABP. Most are based on producing a pseudopotential which reproduces the atomic wavefunctions at distances outside the core region and then allows the pseudowavefunction to go smoothly to zero in the core region. The only input information needed to produce the ABP is the atomic number. Once the ABP is determined, the total energies for the solid can be computed using a density functional approach (Kohn and Sham 1965). The energies of different candidate crystal structures can be compared at varying volumes or pressures, and this can be used to compare the stabilities of the candidate structures with the overall rule that at

**Fig. 4** Total energy calculation for Si as a function of volume (normalized to the atmospheric pressure volume for Si in the diamond structure) for various candidate structural phases. The dashed line illustrates the pressure-induced path for the solid-solid structural phase transition from diamond to β-tin
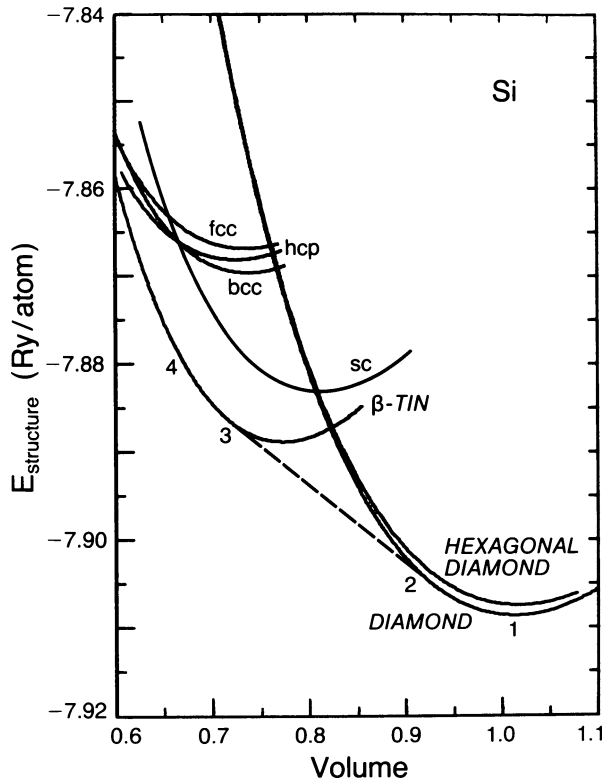


a given volume or pressure, the candidate structure with the lowest energy is the one expected to be chosen by nature. By adding the atomic mass to the input data, it is possible to compute the vibrational properties and the interaction between electrons and vibrations. This scheme allows the determination of crystal structures (see Fig. 4), solid-solid structural phase transitions, lattice constants, elastic constants, bulk moduli, vibrational spectra, electron-phonon couplings, anharmonic properties, and even superconducting properties (Cohen 1982, 1985; Liu and Cohen 1989). The successful predictions of new high-pressure crystalline phases of Si and Ge and the prediction of superconductivity (Chang et al. 1985) in some high-pressure phases of Si are dramatic proofs of the applicability of this ABP-total energy approach for calculating properties of solids.

## 6   Surfaces, Interfaces, and Nanoscience

The discussion above focused on studies of bulk solids, and it was stated that it is possible to use the similar methods for surfaces and interfaces. An important approach for doing the latter and for dealing with localized and lower dimensional systems is based on the concept of a supercell (Cohen et al. 1975). The fundamental

idea is to deal with the lack of periodicity, for example, when a surface is considered, by imposing a different periodicity. For the surface case, by constructing a unit cell with a slab containing the surface atoms plus a region of vacuum above it and repeating this cell infinitely many times, we have the same situation as in the case of a periodic solid with several atoms at a lattice site. Another aspect of this method which is unlike the EPM is that the charge density of the valence electrons which makes up the electronic part of the potential is allowed to readjust to the surface. This is done via a self-consistent method where the calculation is performed for a trial starting arrangement of charge density, and then the output wavefunctions are used to construct a new charge density which is fed back into the calculation. When input and output converge to the same value, then the results are considered self-consistent.

When nanoscience studies became more popular, questions arose as to whether the above methods for bulk three-dimensional configurations would be appropriate for nanoscale systems. Possible concerns involved the nature of the electronic states when there is confinement and different symmetry conditions because of the reduced dimensional nature of the system. The results for calculations of molecular systems, quantum dots, nanotubes, graphene, and a variety of one- and two-dimensional structures proved very successful using a supercell approach. For example, the theoretical methods successfully predicted new nanotube devices; the existence of new nanotubes (e.g., see Fig. 5) containing B, N, and C; the properties of graphene; BN graphene-like layer materials; and superlattices formed from graphene and BN layers. This is a very active area of research, and there is considerable collaboration between theorists, experimentalists, and engineers. It is a field embraced by physicists, chemists, material scientists, computer scientists, device engineers, biologists, etc. because of the size scale which is of interest in all these areas.



**Fig. 5** Schematic illustration of the structure of the predicted BN nanotube subsequently verified by experiment
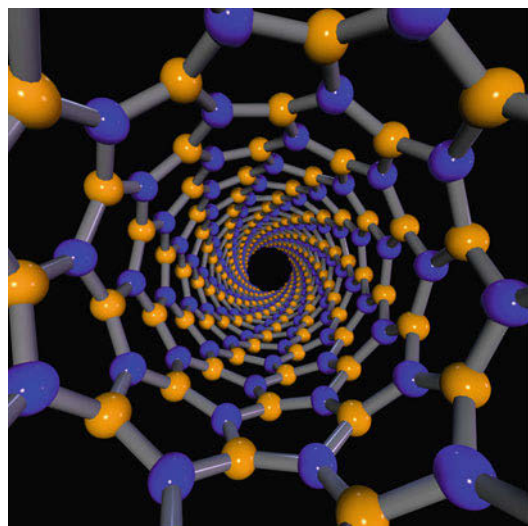
## 7    Emergence

Earlier when describing the models of a solid, it was pointed out that in addition to the interacting atom model, a solid could be described in terms of its elementary excitations. The interacting atom model focuses on the building blocks and their interactions, whereas the elementary excitation model focuses on the responses to probes and the emergent excitations characterizing the solid. In the early part of the twentieth century, there was considerable discussion about reality in science since we do express our findings in terms of human responses. In the elementary excitation model, response functions like resistivity can be described by fictitious particles. For example, the concept of a "hole" which can represent the absence of one electron out of $10^{23}$ is very useful, since it is far easier to explore the dynamics of this positron-like particle than all the remaining electrons after one is removed. So the emergent particle here is a positive fermion with electron characteristics. Similarly, in analogy with the particle-wave duality quantum concept for light, we can treat sound waves in a solid in terms of their associated particles called phonons. The phonon is a boson of spin zero, and many physical measurements involving excitations of sound waves can be treated effectively by using the phonon concept. One may ask whether when using this approach we are losing track of reality. Henri Bergson, a French mathematician, physicist, philosopher, and Nobel Laureate in literature, would have argued that all we know is what we sense; therefore, if this approach explains the observed properties, it is an acceptable view. The concept of emergence and emergent properties is widely discussed not only in condensed matter physics but in many other areas. For example, it has even been proposed that space-time is an emergent property.

Magnetism, superconductivity, and topological insulators with metallic surface currents while having semiconducting bulk properties are examples of emergent properties, so is the quantum Hall effect where resistance can be quantized (see Fig. 6). The fractional quantum Hall effect is an excellent example showing how quasiparticles can emerge from an electronic system having fractional charges like 1/3 e. Since the properties of the systems studied can be explained in terms of these fractionally charged particles, then in the "Bergson sense," this is a fully appropriate way of explaining what the response function is measuring.

For superconductors, the emergence of this intriguing property of solids has excited researchers since 1911. The formulation of the BCS theory (Bardeen et al. 1957) 46 years later appeared to answer all the experimental questions related to superconducting materials discovered up to that time. The central ideas were that the electrons formed pairs – Cooper pairs with zero spin – and these pairs behaved in an almost boson manner allowing zero resistance. However, the pairs were not Heitler-London pairs since for a superconductor like Al, the electrons in a pair are far enough apart to have a million other pairs between them. Roughly speaking, to excite or create a superconducting quasiparticle, one breaks a pair with EM radiation or by raising the temperature. In the BCS scheme, this quasiparticle behaves like a combination of an electron and a hole. The energy needed to break the pair is referred to as a superconducting gap energy which is

## Topological Insulator



Semiconductor in zero magnetic field

--------------------------------------------------------------------------------------------------------------
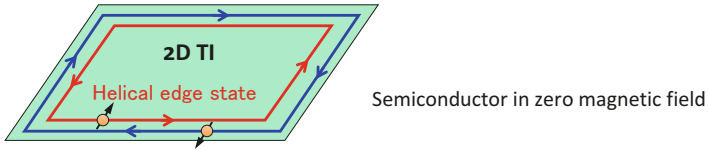
## Quantum Hall effect

Hall bar in a strong magnetic field at low temperatures

$$\sigma_{xy} = ne^2 / h$$

**Fig. 6** Two examples of emergent states are a so-called topological insulator which is a semiconductor having conducting edge states while the bulk remains a semiconductor and the quantum Hall effect where the Hall conductivity is quantized. For the fractional quantum Hall effect, the quasi-electrons can have fractional charges

proportional to the superconducting transition temperature $T_c$. The superconducting gap differs fundamentally from the semiconducting gap. The latter is caused by electron-lattice interactions, while the former is related to the effective electron-electron interactions. The pairing of the electrons is caused by polarizing the lattice which in effect causes an attraction between electrons represented by the electron-phonon pairing parameter $\lambda$. In the simplest form of the BCS theory, if $\lambda$ is larger than the repulsive Coulomb pairing parameter $\mu$, then the electrons can form pairs. Superconductors which can be described by the BCS theory with the attractive pairing interaction originating from electron-lattice (electron-phonon) interactions are considered to be Class 1 superconductors. The properties of these superconductors are well understood, and it is likely that all experiments on these superconductors can be explained using BCS theory.

In the late 1980s, the so-called high $T_c$ superconductors containing copper and oxygen were discovered with $T_c$ values above the liquid nitrogen boiling point. In fact, under pressure, $T_c$ values in the range of 160 K were achieved. However, in the past 30 years, despite extensive and creative research, there is no consensus on a theoretical mechanism for the superconductivity in these systems. A similar situation exists for the Fe-based high $T_c$ superconductors. If we group these and put them into a separate Class 2 category, the general situation presently is that the Class 1 superconductors are completely understood in principle, but there is no general agreement on a theory for Class 2 materials, except perhaps that electron pairs are a central feature here too.

The quest for room temperature superconductivity has been a longtime goal. Because of the high values for $T_c$ of Class 2 superconductors, it was generally felt that the best path to this goal is to try to enhance $T_c$ in this group. However, since a theoretical basis for predicting new Class 2 isn't available, at this point the quest is

in the hands of the experimentalists. In contrast, using BCS theory and the models discussed above, it is possible to predict new Class 1 superconductors (Chang et al. 1985) and explain the properties of materials such as $MgB_2$ (Nagamatsu et al. 2001; Choi et al. 2002) which has the highest reported and experimentally reproduced $T_c$ for Class 1 superconductors. There is a recent report from one group (Drozdov et al. 2015) of superconductivity near room temperature in a hydrogen-sulfur system under pressure, and currently experiments are underway worldwide to reproduce these findings and extend them.

Since Class 1 superconductors are well understood, theory can be used to explore the question of what path is favorable for higher $T_c$ values for Class 1. It is clear that having a large $\lambda$ is an advantage. Also in BCS theory, $T_c$ scales with the phonon frequency. For very large $\lambda$, $T_c \sim \omega \lambda^{1/2}$, where $\omega$ is some average of the phonon frequencies. So we want large $\omega$ and large $\lambda$. It can be shown that $\lambda$, which is dimensionless, can be expressed as the ratio of electronic and lattice spring constants. Hence it is desirable to have strong bonds and light masses, and this makes doped carbon systems desirable. Boron-doped diamond was found to be superconducting first with $T_c \sim 4$ K (Ekimov et al. 2004), and as the doping levels increased, $T_c \sim 11.4$ K. Studies (Moussa and Cohen 2008) of this system predict possible values for $T_c$ of approximately 55–80 K. However, experimentally achieving higher levels of doping with boron is very difficult.

Since the arguments made above and the predictions are more generally applicable than just to diamond, it is important to examine carbon-based materials that can accommodate higher B concentrations. The recently synthesized Q-carbon (Narayan and Bhaumik 2015) is just such a material, and it was found that with a B doping of approximately 17%, this material became superconducting at 36 K (Bhaumik et al. 2017) which is comparable to $MgB_2$ at 39 K (see Fig. 7). At this point, the detailed structural and electronic properties of boron-doped Q-carbon have not been determined. However, many features have been reproduced theoret-



**Fig. 7** Temperature-dependent (normalized) resistivity measurements of B-doped Q-carbon thin films showing the onset of the superconducting transition temperature at 37.8 K. The inset depicts the enlarged view of the superconducting transition showing the transition width to be 1.5 K. (Bhaumik et al. 2017)
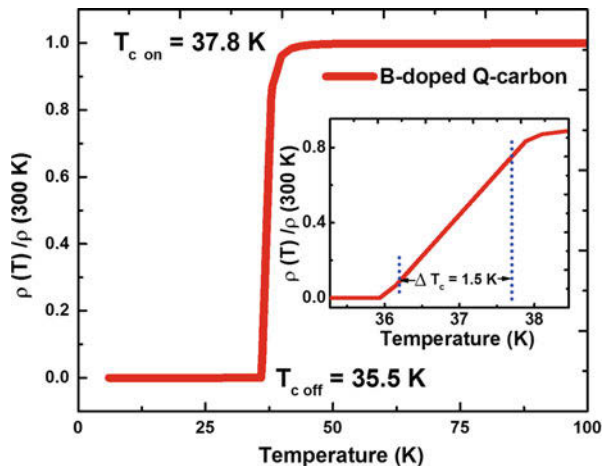
ically (Sakai et al. 2018) using a simulated amorphous carbon system with boron atoms replacing the carbon atoms one by one. It is found that the boron acceptor states can be either shallow or deep, depending on the surrounding geometries. The shallow acceptor states are important for achieving superconductivity because they contribute to the holes at the Fermi energy. Hence this property can be controlled by choosing specific substitutional sites. The electron-phonon coupling is also studied to show that the shallow acceptor states can induce a superconductivity transition as in boron-doped diamond. By calculating the Eliashberg spectral function $\alpha^2 F(\omega)$, $\lambda$ and $T_c$ can be determined. It was found that for a 14% boron doping, $T_c = 37$ K and $\lambda = 1.11$. This result is in agreement with the measurements for B-doped Q-carbon (see Fig. 8). This theoretical approach can be used to calculate other systems and other dopings to maximize $T_c$ for related systems (see Fig. 9). However, it should be noted that at this point, the experimental results of Bhaumik et al. (2017) have not been reproduced by other researchers.
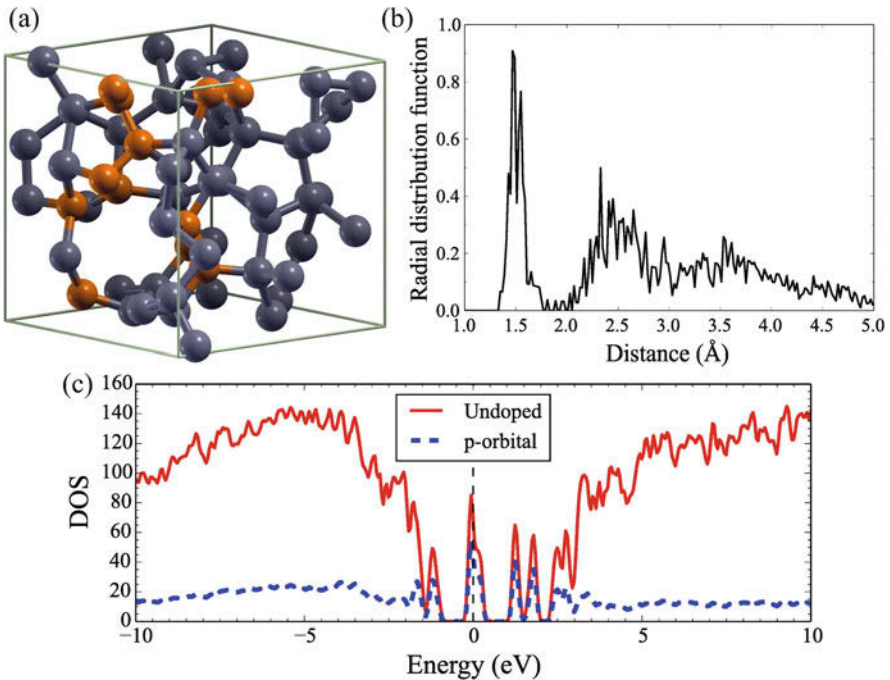


**Fig. 8** (**a**) Ball-and-stick model chosen for amorphous carbon. Orange and gray spheres represent threefold and fourfold coordinated carbon atoms, respectively. (**b**) Radial distribution function of amorphous carbon. (**c**) Density of states (in states/spin/Ry/cell) of undoped amorphous carbon (red solid line) and its projection onto p-orbital of threefold coordinated carbon atoms (blue dashed line). The vertical dashed line at 0 eV indicates the Fermi level. A Gaussian broadening width and energy grid of 0.05 eV is used. (Sakai et al. 2018)

Values of $\eta$

|  | $\eta$ (eV/Å$^2$) | $0.183\sqrt{\lambda\langle\omega^2\rangle}$ (K) | EXP |
|---|---|---|---|
| C (diamond)* | 54 | 290 | ~ 10 |
| C (graphite)* | 48 | 270 | ? |
| BN* | 36 | 240 | ? |
| Si* | 10 | 82 | ~10 |

*at peak of $\eta(E)$

**Fig. 9** Table of the calculated peak values for the energy-dependent electronic spring constant and the corresponding predicted superconducting transition temperatures for four materials. The current experimental values are also shown. (Moussa and Cohen 2006)

## 8    Conclusions

The discussion in this overview of models for solid systems and their evolution over the past century focused on the fact that often the models and theories explaining and predicting properties of solids and in other areas of science start with empirical approaches and develop into schemes that can be classified as being more fundamental or ab initio. In addition to discussing the models and modeling approaches, a few important examples of their applications to recent studies of semiconductors, nanostructures, and superconductors were given. These fields are extremely active at this time motivated by the hope that understanding will lead to new materials which will allow better and smaller devices and higher temperature superconductors. This theme of the importance of motivation from applied science resulting in advances in basic science and vice versa is emphasized here. For example, the quest for a deeper understanding of semiconductors led to new methods for electronic structure calculations which gave us a deeper understanding of the quantum nature of materials. For superconductors, the motivation to understand superconductivity and possibly predict new useful superconductors stimulated fundamental research which led to the development of the BCS theory which was not only a major achievement in condensed matter physics, but it has had an impact on other areas of physics. There is, of course, the influence of applied physics research which led to the development of new instruments such as the laser and the scanning tunneling microscope. New instruments for probing matter and the synthesis of new materials have been essential for the advancement of this area of science. It has been the synergy of pure and applied approaches to the study of matter and its diversity and connections to other fields that has made this area of science and engineering so important.

# References

Ashcroft NW, Mermin ND (1976) Solid state physics. Cambridge University Press, Cambridge, UK

Bardeen J, Cooper LN, Schrieffer JR (1957) Theory of superconductivity. Phys Rev 108:1175

Bhaumik A, Sachan R, Narayan J (2017) High-temperature superconductivity in boron-doped Q-carbon. J ACS Nano 11:5351

Chang KJ, Cohen ML, Mignot JM, Chouteau G, Martinez G (1985) Superconductivity in high-pressure metallic phases of Si. Phys Rev Lett 54:2375

Choi HJ, Roundy D, Sun H, Cohen ML, Louie SG (2002) The origin of the anomalous superconducting properties of $MgB_2$. Nature 418:758

Cohen ML (1982) Pseudopotentials and total energy calculations. Phys Scr T1:5

Cohen ML (1985) Calculation of bulk moduli of diamond and zinc-blende solids. Phys Rev B 32:7988–7991

Cohen ML, Bergstresser TK (1966) Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zincblende structures. Phys Rev 141:789

Cohen ML, Schlüter M, Chelikowsky JR, Louie SG (1975) Self-consistent pseudopotential method for localized configurations: molecules. Phys Rev B 12:5575

Cohen ML, Chelikowsky JR (1988) Electronic structure and optical properties of semiconductors. Springer, Berlin

Cohen ML, Louie SG (2016) Fundamentals of condensed matter physics. Cambridge University Press, Cambridge, UK

Drozdov AP, Eremets MI, Troyan AI, Ksenofontov V, Shylin SI (2015) Conventional superconductivity at 203 Kelvin at high pressures in the sulfur hydride system. Nature 525:73

Ekimov EA, Sidorov VA, Bauer ED, Mel'nik NN, Curro NJ, Thompson JD, Stishov SM (2004) Superconductivity in diamond. Nature 428:542

Fermi E (1934) On the pressure shift of the higher levels of a spectral line series. Nuovo Cimento 11:157

Kittel C (1996) Introduction to solid state physics, 7th edn. Wiley, New York

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133

Liu AY, Cohen ML (1989) Prediction of new low compressibility solids. Science 245:841

Moussa JE, Cohen ML (2006) Two bounds on the maximum phonon-mediated superconducting transition temperature. Phys Rev B 74:094520

Moussa JE, Cohen ML (2008) Constraints on Tc for superconductivity in heavily boron-doped diamond. Phys Rev B 77:064518

Nagamatsu J, Nakagawa N, Muranaka T, Zenitani Y, Akimitsu J (2001) Superconductivity at 39 K in magnesium diboride. Nature 410:63

Narayan J, Bhaumik A (2015) Novel phase of carbon, ferromagnetism, and conversion into diamond. J Appl Phys 118:215303

Phillips JC, Kleinman L (1959) New method for calculating wave functions in crystals and molecules. Phys Rev 116:287–294

Phillips JC (1973) Bonds and bands in semiconductors. Academic, New York

Sakai Y, Chelikowsky JR, Cohen ML (2018) Simulating the effect of boron doping in superconducting carbon. Phys. Rev. B 97:054501

Walter JP, Cohen ML (1971) Electronic charge densities in semiconductors. Phys Rev Lett 26:17–19

Yu PY, Cardona M (1996) Fundamentals of semiconductors. Springer, Berlin

# The Long and Winding Road: Predicting Materials Properties Through Theory and Computation

**4**

Giulia Galli

## Contents

**Abstract**

First-principles simulations of materials provide both computational microscopes and predictive tools, which we aspire to turn into design strategies for materials with target properties. One requisite to meet this goal is the enablement of predictions of material properties on a large scale, so as to generate a vast amount of validated computational data that may eventually be used to solve inverse problems. However it is challenging to use big data to address the "why question." First-principles calculations of specific materials and properties can instead be extremely effective at answering the "why question," namely, at unraveling mechanisms and providing fundamental, physical insights, thus paving the way to innovative design strategies. In this chapter, we present first-principles predictions of material properties relevant to energy conversion processes. We also discuss some open challenges related to automated integration of theory and computation with experiments and with validated, interpreted data.

G. Galli (✉)
Department of Chemistry, Institute for Molecular Engineering, The University of Chicago, Chicago, IL, USA
e-mail: gagalli@uchicago.edu

# 1    Introduction

Science and engineering of materials is a vast field ultimately enabling the development of new technologies, with impact in energy, quantum information science, medicine and health, and national security. Almost 10 years ago, in the United States, President Obama launched the Materials Genome Initiative (MGI) (Obama 2011), with a speech at Carnegie Mellon (June 2011). MGI has been a game changer in the way the scientific community thinks of and approaches research in materials. The ambitious goal set by Obama, "To help businesses discover, develop, and deploy new materials twice as fast . . . ," catapulted research on materials at the forefront of science and engineering on the national scene and pointed out clearly and forcefully that business as usual was not an option in materials research. The MGI pushed toward innovation, to developing brand new techniques to make, study, and predict materials, and recognized theory and computation not only as an integral part of the innovative process but as a driver seat player. Projects in developing predictive tools and databases for materials flourished, for example, the Materials Projects (Jain et al. 2013a) initiated at MIT and then established at LBNL and several related initiatives (Curtarolo et al. 2012; Jain et al. 2013a, 2016; Saal et al. 2013; Bhat et al. 2015; Kalidindi and De Graef 2015; Rajan 2015; Blaiszik et al. 2016; Thygesen and Jacobsen 2016; Chard et al. 2018).

In 2015 the US DOE established for the first time computational materials centers (CMS) to develop methods and software to predict materials properties – importantly to develop software open to the community – thus further enhancing the pace of research and innovation. Three centers were established, two at National Laboratories (BNL: https://www.bnl.gov/comscope/ and ANL: http://miccom-center.org/) and one on a university campus (USC: https://magics.usc.edu/). A year later two additional centers were created, at LBNL (http://c2sepem.lbl.gov/) and ORNL (https://cpsfm.ornl.gov/). These centers were born in the ecosystems of the energy hubs conceived by Steven Chu (JCESR (www.jcesr.org) and JCAP (https://solarfuelshub.org/)); the hubs have a clear mission toward a societal grand challenge, climate change (global warming), and within that ecosystems, many of the projects of DOE centers and other agencies, notably NSF (www.nsf.gov/pubs/2019/nsf19516/nsf19516.htm), focused their research on functional materials for energy.

A question that arises in many instances when discussing the impact of MGI and the US CMS, as well similar centers established in in Europe and Asia, is simply what's new relative to the deployment of software in the semiconductor and pharmaceutical industry: in these industries codes for materials and molecular systems have been used for decades. However such codes have traditionally been used as mostly end of the line engineering tools, e.g., to test molecules that would not be synthetized first in the laboratory or to help design chips that had already been planned, based on specific material choices. The theories, codes, and software developed and pushed by MGI-like ideas are meant to become (and in some cases are already becoming) beginning of the line multi-scale tools to produce innovative

ideas on materials that have not yet been made, have not yet been planned, and do not even come from conventional synthetic and fabrication sources.

These theories and codes are envisioned to be general enough to meet two upcoming scientific revolutions: artificial intelligence (AI) and big data (De Mauro et al. 2015) and quantum information science and technology (Bennett and DiVincenzo 2000). It is important to emphasize that *general* applicability is a key, distinctive feature relative to the codes used in the semiconductor industry, for example, which have traditionally been developed for specific tasks, targeting one specific class of materials and processes. General codes may be used to produce the data needed for AI technologies for vast classes of materials. Computer generated data may then be part of design strategies that include innovative feedback loops with experiments, as well as strategies to make data reproducible and available to the scientific community worldwide (see, e.g., Govoni et al. 2019).

Contemporary computational methods and database mining techniques have already made tremendous strides in the prediction of equilibrium properties of materials that exhibit simple morphologies. However, the functionality of modern materials depends critically on the integration of dissimilar components and on the interfaces that arise between them. Hence the atomic- and molecular-scale manipulation of these components and the heterogeneous structures that emerge from them are key to materials design. In particular, the controlled and driven assembly of building blocks into hierarchical systems, as well as the control of defects and complex morphologies, offers the opportunity to create artificial materials that do not exist in nature and that exhibit superior physical properties for, e.g., emerging energy and quantum information technologies.

The simulations of heterogeneous materials and of the assembly process of artificial materials are much less advanced than the study of equilibrium properties. In order to accelerate the discovery of innovative functional materials, it will be key to acquire the ability not only to compute the properties of the end product but also to simulate and validate the assembly processes that take place during synthesis and fabrication. In addition, in order to design materials relevant to many technologies, it is essential to predict functionalities of systems with complex defective structures and ultimately complex morphologies and to simulate and eventually engineer the basic mass, charge, and energy transport phenomena, as pictorially illustrated in Fig. 1. We emphasize that most transport phenomena, e.g., electron transport, and phenomena involved in the spectroscopic characterization of materials, which involve interaction with light, are inherently quantum mechanical and thus require a first-principles, quantum mechanical treatment of interatomic interactions, at the atomistic scale.

In the following, we focus on two examples (interfaces for energy conversion processes, in Sect. 2 and materials composed of complex building blocks, in Sect. 3), and we describe recent progress in describing heterogeneous, defective materials with complicated morphologies using first-principles methods (Martin 2004, Martin et al. 2016). We aim at showing the importance of unraveling mechanisms and providing fundamental, physical insights, in order to pave the way to material design

**Fig. 1** Integrated predictions of multiple properties are key to define effective design strategies for materials with target characteristics. These properties encompass the atomistic structure of the material, possibly derived from the assembly of complex building blocks, the response to electromagnetic fields (light) used to probe and characterize the material and transport properties, including mass, charge, and heat transport
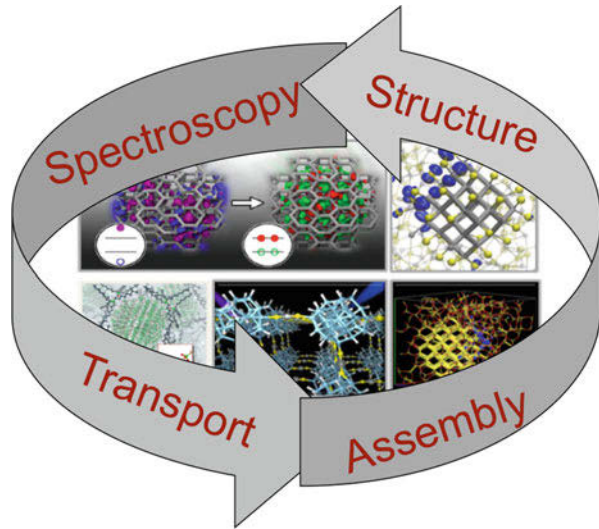


strategies. We close (Sect. 4) by describing open challenges in understanding and predicting synthetic pathways to obtain materials with target properties.

## 2  Energy Conversion at Interfaces from First Principles

In Figs. 2 and 3, we show some of the key processes and properties that one aims at understanding to establish a structure-function relationship and eventually predict optimal materials for solar-to-fuel and solar energy conversion, respectively. The figures illustrate the complexity of the predictive endeavor and the multitude of properties one should be able to compute, validate, and ultimately integrate with experiments. We concentrate here on materials for photo-electrochemical cells (PECs; Fig. 2).

The generation of hydrogen from water and sunlight through PECs is one of the promising approaches investigated by the scientific community in the last decades for producing sustainable carbon-free energy (Walter et al. 2010; McKone et al. 2013; Pham et al. 2017). A key aspect to building an efficient PEC is the availability of Earth-abundant semiconducting photoelectrode materials that can absorb sunlight and eventually drive water-splitting reactions when interfaced with the liquid. Despite steady efforts, no single material has yet been found that simultaneously satisfies the efficiency and stability required for the widespread commercialization of hydrogen technology, and efforts have been concentrated on architectures composed of different materials, notably absorber solids interfaced with catalysts. Hence, understanding the properties of the interfaces between the various components is key to predict novel systems and eventually to optimize the device performance.
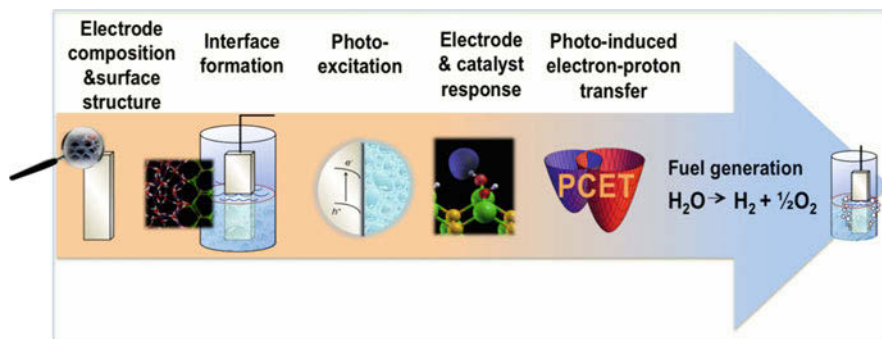
**Fig. 2** Pictorial representation of key processes and systems involved in water-splitting reactions occurring on a catalytic surface, which starts with harvesting light to form charge carriers and involves proton-coupled electron transfer (PCET) processes
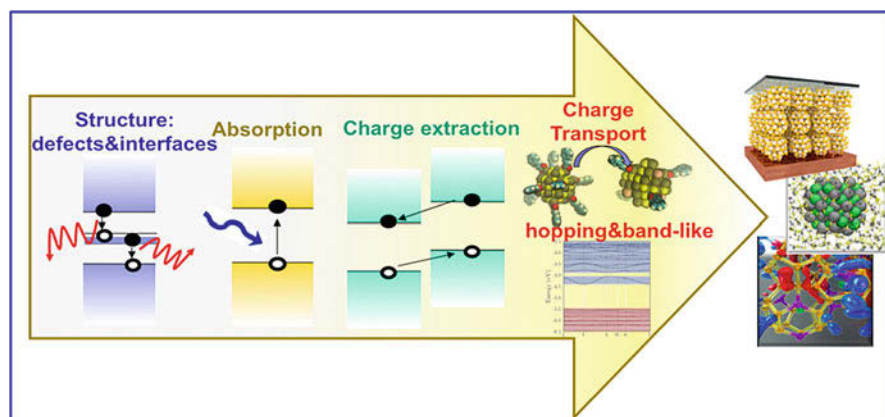


**Fig. 3** Pictorial representation of the key physical processes involved in the prediction and design of nanostructured semiconducting materials for solar energy conversion, including ensembles of nanoparticles (NPs), embedded NPs, and inorganic clathrates (icons on right hand side). Electronic (absorption, photoemission, and band offsets) and transport properties may be obtained from calculations based on density functional and many-body perturbation theory

In this regard, electronic and structural properties of absorbers/catalysts/water interfaces play a critical role, as rapid charge transfer between the photoelectrode, the catalysts, protective layers, and electrolytes is required for efficient fuel production. Interfacial structural and electronic properties of PECs are of course intertwined. For example, band edge positions of photoelectrode absorbers depend on the surface termination, the reconstruction, and the concentration of impurities and defects. In addition, the stability of the absorbers against oxidation (reduction) is determined by the relative energy between their valence band maximum (conduction band minimum) and intrinsic oxidation (reduction) potential. Such a complex interplay results in a multi-property optimization problem which, given recent

advances in high-performance computing and sophisticated electronic structure theories and codes (Kresse and Hafner 1993; Soler et al. 2002; Gygi 2008; Blum et al. 2009; Giannozzi et al. 2009; Hutter et al. 2014; VASP, Kresse and Furthmüller (1996a, b) Kresse and Hafner (1994) www.vasp.at; SIESTA, www.icmab.es/siesta; Qbox, www.qbox-code.org; FHI-AIMS, http://aims.fhi-berlin.mpg.de/; Quantum Espresso, www.quantum-espresso.org; CP2K, www.cp2k.org; CPMD, www.cpmd.org;), is now conceivable to tackle using first-principles simulations.

In recent years, it has been successfully demonstrated that first-principles calculations can be employed to scan thousands of combinations of elements across the entire periodic table to suggest new photoelectrode candidates (Greeley et al. 2006; Jain et al. 2013b; Castelli et al. 2015). However, computational screening schemes available thus far in the literature have mostly focused on bulk properties of candidate materials, and only recently the structural and chemical properties of surfaces and interfaces with the electrolyte have attracted the attention that they deserve to build successful design strategies. To paraphrase what Herbert Kroemer so elegantly pointed out is his Nobel lecture (Kroemer 2000) on semiconductor heterojunctions, the interface is *still* the device! As shown by us and others, the effective predictions of band offsets for water photocatalysis require the simulations of the electronic structure of solvated surfaces at finite temperature and in the case of oxide surface, importantly of *defective* solvated surfaces (Gerosa et al. 2018).

For example, in a case we have recently studied, $WO_3$ (Gerosa et al. 2018 and reference therein), we have shown that the average potential energy difference at the interface of pristine and defective $WO_3$ varies by $\sim$1 eV and that solvation is absolutely critical (see Fig. 4). In addition, we have shown the key importance of using a high level of theory, beyond the widely used density functional theory (DFT)
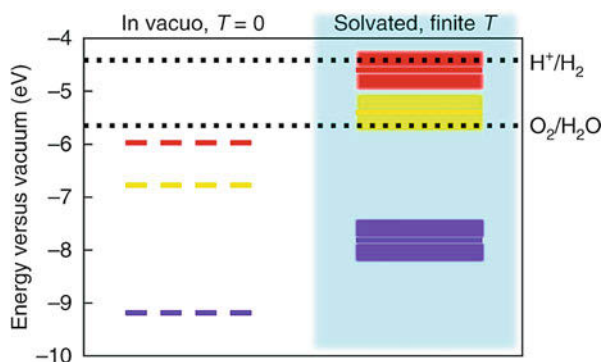


**Fig. 4** Energy levels (valence band maximum, blue; conduction band minimum, red; defect state due to oxygen vacancies; yellow) of a $WO_3$ surface in vacuo, at $T = 0$ and at room temperature, in the presence of water (solvated). The energy levels have been obtained using first-principles molecular dynamics simulations and calculations at the many-body perturbation theory level (GW), starting from electronic states computed with hybrid density functionals (From Gerosa et al. 2018). Note the striking difference of the positions of the levels on the right and left hand side, relative to the redox levels of liquid water

(Hohnberg and Kohn 1964; Kohn and Sham 1965; Martin 2004) and hybrid DFT (Perdew et al. 1996 ; Heyd et al. 2003, 2006) to carry out predictive calculations. The latter have allowed us to understand that the excess charge present at defective $WO_3$ surfaces due to oxygen vacancies forms a large 2D polarons ($\sim$10 A radius) on the plane of the surface; the predicted charge localization properties hint at possible formation of stable ($OH^-$) groups at the surface in contact with water and at the fact that holes transferred to water would then form a highly reactive $(OH)^*$, a possible precursor of water-splitting reactions. Altogether our calculations have identified three major factors determining the chemical reactivity of oxide absorbers interfaced with water: the presence of surface defects, the dynamics of excess charge at the surface, and finite temperature fluctuations of the surface electronic orbitals. These general descriptors are essential for the understanding and prediction of optimal oxides for water oxidation.

This was presented as an example of the importance of gaining fundamental physical insight into descriptors in order to define material design strategies and in particular into non-intrinsic properties of materials such as interfaces between complex components and defects present at finite temperature. We now turn to a second example of materials made of complex, nanostructured building blocks, where again interfaces – specifically buried interfaces – dominate the scene.

## 3   Building Blocks for Electronic Materials and Materials for Energy Conversion

In this section we consider materials made of nanostructured building blocks, in particular semiconducting colloidal nanocrystals (NCs) (Scalise et al. 2018; Greenwood et al. 2018; Talapin et al. 2010). Systems built from the assembly of these "artificial atoms" are emerging as tunable, earth-abundant, and potentially nontoxic materials for solar energy conversion, light emission, and electronic applications (Talapin 2012; Kovalenko 2013; Wippermann et al. 2013, 2014, 2016). The electronic and transport properties of NC-based solids depend on many factors that encompass the intrinsic characteristics of the individual NCs, for example, their shape, size, and composition, as well as their surface chemistry and mutual interactions. Organic ligands traditionally used in NC synthesis play a central role in controlling shape and size, as well as in driving self-assembly into superlattices. However, these ligands are often composed of long hydrocarbon chains, which create an insulating barrier that leads to low charge carrier mobilities. Significantly higher mobilities could be achieved by using inorganic ligands, and their use has enabled significant performance improvements of NC-based solar cells, transistors, and lasers.

For example, InAs and CdSe NCs capped with molecular metal chalcogenide complexes (MCC) were shown to exhibit high electron mobilities (Lee et al. 2011; Liu et al. 2013), with III–V-based nanomaterials preferable for commercial applications due to their lower toxicity. However, the atomistic structure of these materials is difficult to characterize, in particular that of the NC surfaces and
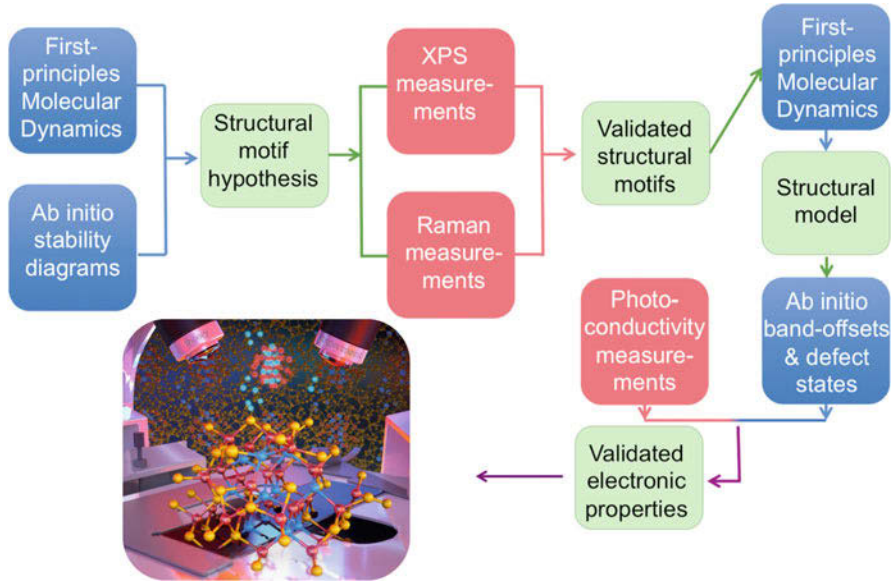
**Fig. 5** Schematic representation of the integrated experimental and computational strategy adopted to obtain validated structural models and electronic properties of all inorganic semiconductor materials composed of colloidal nanocrystals, represented in the inset on the bottom right (see text)

interfaces, whose control is required to engineer systems with the desired properties. Recently we proposed (Scalise et al. 2018) a strategy to model a broad class of nanocrystal-in-glass systems that extends significantly beyond semiconductor quantum dots and MCC ligands.

Our strategy is summarized in Fig. 5. By combining first-principles molecular dynamics (MD) and ab initio stability diagram calculations (ab initio electronic structure calculations of surface energies and stability), main structural motifs were identified; in particular the structure of buried interfaces was determined. Before proceeding to derive a complete structural model, the motifs obtained computationally were experimentally validated, by carrying out XPS and Raman measurements, which both confirmed the results of the calculations. Using these validated structural motifs as a starting point of additional first-principles MD simulations, a structural model consistent with experiment was finally derived and used to analyze the electronic structure of the composite material. The predicted electronic states were used to interpret and understand the reasons for the measured negative photoconductivity, thus identifying specific reasons giving rise to properties that had remained unexplained and controversial for some time.

Overall, by combining electronic structure calculations and first-principles molecular dynamics (MD) simulations with experiments, we showed that the ligands are not absorbed as intact units but rather they decompose on contact with

the NC surface to form an amorphous matrix that encapsulates the nanoparticles (NPs). The intrinsic electronic properties of the isolated NCs are greatly modified in the matrix, whose atomistic structure plays a key role in enabling an efficient electronic transport. The structural model derived in this way permitted an explanation of the origin of the measured negative photoconductivity of the nanocomposite. This was presented as an example of novel material properties emerging when assembling building blocks at the nanoscale and as an example of the importance of tightly integrating theory, computation, and experiments. The future challenge will be to achieve such integration automatically and to define general validation strategies appropriate for broad classes of systems.

## 4   The Synthetic Challenge

One of the open challenges in computational materials science is the understanding and prediction of how to synthesize materials with target properties (De Yoreo et al. 2016). Using experimental data and simulations, the challenge is to establish correlations between synthesis protocols (SP) and material structure (M) and between synthesis protocols and material properties (P). This endeavor requires the solution of both direct and so-called inverse problems (Kaipio and Somersalo 2005). The former include answering first the question: Given a synthesis protocol (SP), what material (M) does one obtain? This forward problem is a grand challenge for predictive, computational methods. We still lack well-defined physics models that can describe synthesis; in addition "realistic" materials encompass complex descriptors including crystal structure, morphology, defects, and surface coverings. An even more complex forward problem concerns materials properties and the following questions: Given a synthesis protocol, which properties (P) does one get? This forward problem is clearly not unrelated to the first one (SP $\rightarrow$ M); however, it poses additional experimental and theoretical challenges. In particular, on the theory side, the prediction of certain complex properties is still in the making, e.g., obtaining optoelectronic and vibrational spectroscopic data and transport data, which require the use of sophisticated, cannot yet be used efficiently to acquire large amount of data for broad classes of systems.

The ultimate goal of the science of synthesis research is to solve the inverse problems associated with the forward problems mentioned above: Given a desired material, what synthesis protocol should be used to obtain it? Given a set of desired materials properties, what synthesis protocol should be used to obtain them? Solutions to inverse problems are found by solving many forward problems in a regression loop, which requires forward problems to be rapidly computable and the use of data analytic approaches, designed to mine experimental and theoretical data.

Ultimately, by seeking correlations between data and synthesis conditions, one will enable the discovery of materials and synthesis ontologies, i.e., a set of descriptors linking materials and synthetic pathways (e.g., crystal phase and synthesis temperature). Accurate ontologies are not yet known for synthesis, and their definition may come from the use of machine learning (ML) to search for the

most relevant descriptors for a given outcome and to relate different descriptors used by different researchers. We expect that with the increasing population of experimental and computational databases in the materials science community, ML workflows (Jain et al. 2015; Pizzi et al. 2016; Meng and Thain 2017; Adorf et al. 2018; Freire and Chirigati 2018) may be used to train models of materials synthesizability and properties and hence to predict novel materials.

In closing, we would like to comment on data and data availability. A key guiding principle for materials research is to make data findable, accessible, interoperable, and reusable. The reproducibility of experiments and computations and of the corresponding results is an important and critical part of the overall research process of all scientific disciplines and in particular of materials predictions heavily relying on large amount of data. Yet the data presented in most published scientific papers are not made available to the community, and the procedures followed to obtain or generate the data are often not articulated step by step or in any detail. Hence making all data available to the public (Govoni et al. 2019), on a paper-by-paper basis, so as to increase experimental and computational rigor in reporting results, together with transparency, should become integral part of the research process. This endeavor will also greatly contribute to devising improved validation procedures for computational data as well as establishing experimental and computational automatic feedback loops.

# References

Adorf CS, Dodd PM, Ramasubrami V, Glotzer SC (2018) Simple data and workflow management with the signac framework. Comput Mater Sci 146:220

Bennett CH, DiVincenzo DP (2000) Quantum information and computation. Nature 404:247–255

Bhat TN, Bartolo LM, Kattner UR, Campbell CE, Elliot JT (2015) Strategy for extensible, evolving terminology for the materials genome initiative efforts. JOM 67:1866

Blaiszik B, Chard K, Pruyne J, Ananthakrishnan R, Tuecke S, Foster I (2016) The materials data facility: data services to advance materials science research. JOM 68:8

Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M (2009) Ab initio molecular simulations with numeric atom-centered orbitals. Comput Phys Commun 180:2175–2196

Castelli IE, Hüser F, Pandey M, Li H, Thygesen KS, Seger B, Jain A, Persson KA, Ceder G, Jacobsen KW (2015) New light-harvesting materials using accurate and efficient bandgap calculations. Adv Energy Mater 5:1400915

Chard K, Dart E, Foster I, Shifflett D, Tuecke S, Williams J (2018) The Modern Research Data Portal: a design pattern for networked, data-intensive science. PeerJ Comput Sci 4:e144

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, Nelson LJ, Hart GLW, Sanvito S, Buongiorno-Nardelli M, Mingo N, Levy O (2012) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. Comput Mater Sci 58:227

De Mauro A, Greco M, Grimaldi M (2015) What is big data? A consensual definition and a review of key research topics. AIP Conf Proc 1644:97–104

De Yoreo J, Mandrus D, Soderholm L (2016) Basic research needs for synthesis science: report of the basic energy sciences workshop on synthesis science for energy relevant technology. https://science.energy.gov/~/media/bes/pdf/reports/2017/BRN_SS_Rpt_web.pdf

Freire J, Chirigati F (2018) Provenance and the different flavors of computational reproducibility. IEEE Data Eng Bull 41(1):15–26. http://sites.computer.org/debull/A18mar/p15.pdf

Gerosa M, Gygi F, Govoni M, Galli G (2018) The role of defects and excess surface charges at finite temperature for optimizing oxide photoabsorbers. Nat Mater 17:1122–1127

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) QUANTUM ESPRESSO: a modular and open- source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Govoni M, Munakami M, Tanikanti A, Skone JH, Runesha HB, Giberti F, de Pablo J, Galli G (2019) Qresp, a tool for curating, discovering, and exploring reproducible scientific papers. Sci Data 6:190002. https://www.nature.com/articles/sdata20192

Greeley J, Jaramillo TF, Bonde J, Chorkendorff IB, Nørskov JK (2006) Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. Nat Mater 5:909–913

Greenwood AR, Vörös M, Giberti F, Galli G (2018) Emergent electronic and dielectric properties of interacting nanoparticles at finite temperature. Nano Lett 18:255–261

Gygi F (2008) Architecture of Qbox: a scalable first-principles molecular dynamics code. IBM J Res Dev 52:137–144

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207

Heyd J, Scuseria GE, Ernzerhof M (2006) Erratum: hybrid functionals based on a screened Coulomb potential. J Chem Phys 124:219906

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864

Hutter J, Iannuzzi M, Schiffmann F, VandeVondele J (2014) CP2K: atomistic simulations of condensed matter systems. Wiley Interdiscip Rev Comput Mol Sci 4:15–25

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013a) Commentary: the materials project: a materials genome approach to accelerating materials innovation. Appl Mater 1:011002

Jain A, Castelli IE, Hautier G, Bailey DH, Jacobsen KW (2013b) Performance of genetic algorithms in search for water splitting perovskites. J Mater Sci 48:6519–6534

Jain A, Ong SP, Chen W, Medasani B, Qu X, Kocher M, Brafman M, Petretto G, Rignanese G-M, Hautier G, Gunter D, Persson KA (2015) FireWorks: a dynamic workflow system designed for high-throughput applications. Concurr Comput Pract Exp 27:5037

Jain A, Persson KA, Ceder G (2016) Research update: the materials genome initiative: data sharing and the impact of collaborative ab initio databases. APL Mater 4:053102

Kaipio J, Somersalo E (2005) Statistical and computational inverse problems. Springer, New York

Kalidindi SR, De Graef M (2015) Materials data science: current status and future outlook. Annu Rev Mater Res 45:171. https://www.annualreviews.org/doi/full/10.1146/annurev-matsci-070214-020844

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133

Kovalenko MV (2013) Chemical design of nanocrystal solids. Chimia 67:316–321

Kresse G, Furthmüller J (1996a) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6:15. https://www.sciencedirect.com/science/article/pii/0927025696000080?via%3Dihub

Kresse G, Furthmüller J (1996b) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169. https://www.sciencedirect.com/science/article/pii/0927025696000080?via%3Dihub

Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. Phys Rev B 47:558. https://journals.aps.org/prb/abstract/10.1103/PhysRevB.47.558

Kresse G, Hafner J (1994) Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. Phys Rev B 49:14251. https://journals.aps.org/prb/abstract/10.1103/PhysRevB.49.14251

Kroemer H (2000) Quasi-electric fields and band offsets: teaching electrons new tricks. Nobel Prize Lecture, Stockholm. https://www.nobelprize.org/prizes/physics/2000/kroemer/lecture/

Lee J-S, Kovalenko MV, Huang J, Chung DS, Talapin DV (2011) Band-like transport, high electron mobility and high photoconductivity in all-inorganic nanocrystal arrays. Nat Nanotechnol 6:348–352

Liu W, Lee J-S, Talapin DV (2013) III–V Nanocrystals capped with molecular metal chalcogenide ligands: high electron mobility and ambipolar photoresponse. J Am Chem Soc 135:1349–1357

Martin RM (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, Cambridge

Martin RM, Reining L, Ceperley DM (2016) Interacting electrons: theory and computational approaches. Cambridge University Press, Cambridge

McKone JR, Lewis NS, Gray HB (2013) Will solar-driven water-splitting devices see the light of day? Chem Mater 26:407–414

Meng H, Thain D (2017) Facilitating the reproducibility of scientific workflows with execution environment specifications. Proc Comput Sci 108:705. https://www.sciencedirect.com/science/article/pii/S1877050917306816

Obama B (2011) Remarks by the President at Carnegie Mellon University's National Robotics Engineering Center. The Obama White House Archives. https://obamawhitehouse.archives.gov/the-press-office/2011/06/24/remarks-president-carnegie-mellon-universitys-national-robotics-engineer

Perdew JP, Ernzerhof M, Burke K (1996) Rationale for mixing exact exchange with density functional approximations. J Chem Phys 105:9982

Pham TA, Ping Y, Galli G (2017) Modelling heterogeneous interfaces for solar water splitting. Nat Mater 16:401–408

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci 111:218

Rajan K (2015) Materials informatics: the materials "gene" and big data. Annu Rev Mater Res 45:153. https://www.annualreviews.org/doi/full/10.1146/annurev-matsci-070214-021132

Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). JOM 65:1501

Scalise E, Srivastava V, Janke E, Talapin DV, Galli G, Wippermann S (2018) Surface chemistry and buried interfaces in all-inorganic nanocrystalline solids. Nat Nanotechnol 13:841–848

Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, Sánchez-Portal D (2002) The SIESTA method for ab initio order-N materials simulation. J Phys Condens Matter 14:2745

Talapin DV (2012) Nanocrystal solids: a modular approach to materials design. MRS Bull 37:63–71

Talapin DV, Lee J-S, Kovalenko MV, Shevchenko EV (2010) Prospects of colloidal nanocrystals for electronic and optoelectronic applications. Chem Rev 110:389–458

Thygesen KS, Jacobsen KW (2016) Making the most of materials computations. Science 354:180. http://science.sciencemag.org/content/354/6309/180

Walter MG, Warren EL, McKone JR, Boettcher SW, Mi Q, Santori EA, Lewis NS (2010) Solar water splitting cells. Chem Rev 110:6446–6473

Wippermann S, Vörös M, Rocca D, Gali A, Zimanyi G, Galli G (2013) High-pressure core structures of Si nanoparticles for solar energy conversion. Phys Rev Lett 110:046804

Wippermann S, Vörös M, Gali A, Gygi F, Zimanyi G, Galli G (2014) Solar nanocomposites with complementary charge extraction pathways for electrons and holes: Si embedded in ZnS. Phys Rev Lett 112:106801

Wippermann S, He Y, Vörös M, Galli G (2016) Novel silicon phases and nanostructures for solar energy conversion. Appl Phys Rev 3:040807

# Big Data-Driven Materials Science and Its FAIR Data Infrastructure

**5**

Claudia Draxl and Matthias Scheffler

## Contents

**Abstract**

This chapter addresses the fourth paradigm of materials research – big data-driven materials science. Its concepts and state of the art are described, and its challenges and chances are discussed. For furthering the field, open data and an all-embracing sharing, an efficient data infrastructure, and the rich ecosystem of computer codes used in the community are of critical importance. For shaping this fourth paradigm and contributing to the development or discovery of improved and novel materials, data must be what is now called FAIR – *Findable*, *Accessible*, *Interoperable*, and *Repurposable/Reusable*. This sets the stage for advances of methods from artificial intelligence that operate on large data sets to find trends and patterns that cannot be obtained from individual calculations and not even directly from high-throughput studies. Recent progress is reviewed and

C. Draxl (✉)
Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: claudia.draxl@physik.hu-berlin.de

M. Scheffler
Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany
e-mail: scheffler@FHI-Berlin.mpg.de

           49

demonstrated, and the chapter is concluded by a forward-looking perspective, addressing important not yet solved challenges.

## 1 Introduction

Materials science is entering an era where the growth of data from experiments and computations is expanding beyond a level that can be handled by established methods. The so-called 4 V challenge – concerning *Volume* (the amount of data), *Variety* (the heterogeneity of form and meaning of data), *Velocity* (the rate at which data may change or new data arrive), and *Veracity* (the uncertainty of data quality) – is clearly becoming eminent. Most importantly, however, is that big data of materials science offer novel, extraordinary, and expansive opportunities for achieving scientific knowledge and insight. These opportunities require new research concepts and lines of thought. While this chapter focuses on computational materials science, we emphasize that what is described here applies to experimental data as well.

Today's common approach in computational materials science is to publish results as focused research studies, reporting only those few data that are directly relevant for the respective topic. Thus, even from very extensive computations (expending thousands or millions of CPU core hours), very few results are shared with the community. Most data, in particular when they were not deemed of immediate relevance, are kept private or even thrown away. Since a few years, the community of computational materials science and engineering is undergoing "a change in scientific culture" and has started the extensive sharing of data of this field. Sharing of *all* data, i.e., the full input and output files of computations, implies that calculations don't need to be repeated again and again, and the field will have access to big data which can be used in a totally new research manner, i.e., by artificial intelligence methods (Draxl and Scheffler 2019 and references therein). As will be elaborated in the next sections, one can find structure and patterns in big data, gaining new insight that cannot be obtained by studying small data sets, and in this way even allegedly inaccurate data can get value. A popular example from daily life of the impact of big data analytics is the tracking of the movement of mobile phones which provides instantaneous information on traffic flow and jam. Another example is the local information of google searches for *flu symptoms and medicine* which reflect the spreading of a flue wave. The latter example also illustrates the danger of data analytics, as the "google flue trend" worked well for the first few years, but then, in 2012, some aspects became unreliable. Reasons may be, among others, anxiety-stimulated searches caused by reports in public and social media and changes in google search technology, e.g., recommending searches based on what others have searched (Lazer et al. 2014). This example also illustrates that big data should not be viewed as substitute but as complement to traditional analysis. Specifically, we note that even if the amount of data may be big, the independent information may be small when data are correlated, or data may be even irrelevant

or misleading for the purpose of interest. If such aspects are not considered properly, a statistical analysis will be flawed.

Overcoming the "silo mentality" of computational materials science and the development and implementation of concepts for an extensive data sharing was achieved by the NOMAD Center of Excellence (NOMAD; Draxl and Scheffler 2019), considering all aspects of what is now called the FAIR data principles (NOMAD 2014, Wilkinson et al. 2016): Data are *Findable* for anyone interested; they are stored in a way that make them easily *Accessible*; their representation follows accepted standards (Ghiringhelli et al. 2016, 2017a); and all specifications are open – hence data are *Interoperable*. All of this enables the data to be used for research questions that could be different from their original purpose; hence data are *Repurposable*. We note that the NOMAD CoE uses the term repurposable, while in the FAIR concept it was termed reusable. The former makes explicit that data can not only be reused but even used for a different purpose. Obviously, FAIR data also become usable for artificial-intelligence tools.

The chapter is structured as follows. In Sect. 2, we briefly summarize the history of the four research paradigms of materials science, with particular focus on the fourth one, "big data-driven materials science." Section 3 then stresses the importance of an extensive data sharing for the advancements of science and engineering. Section 4 addresses artificial-intelligence concepts for materials science data with some specific examples. Finally, in Sect. 5, we give an extensive outlook on the developments and open questions of big data-driven materials science.

## 2 The Four Research Paradigms of Material Sciences

The historical evolution of methods and concepts of materials science are sketched in Fig. 1. We recall that experimental research dates back to the Stone Age, and the basic techniques of metallurgy were developed in the Copper and Bronze Ages which started in the late sixth millennium BCE. The control of fire prompted a major experimental breakthrough. Toward the end of the sixteenth century, scientists began to describe physical relationships through equations. Well-known names from the early days are Tycho Brahe (1546–1601), Tomas Harriot (ca. 1560–1621), Galileo Galilei (1564–1642), Johannes Kepler (1571–1630), Isaac Newton (1643–1727), and Gottfried Wilhelm Leibniz (1646–1716). The latter two also developed the concept of the mathematical differential and derivatives. Thus, analytical equations became the central instrument of theoretical physics. Second from the left in Fig. 1, this new way of thinking – the 2. paradigm – is symbolized by the Schrödinger equation. Needless to say, the first paradigm, the empirical and experimental sciences, did not become obsolete, but theoretical physics represents an important complementary methodology.

Since the 1950s electronic structure theory was advanced for materials by J. C. Slater (e.g., Slater 1937, 1953, 1965, 1967; Slater and Johnson 1972), the Monte
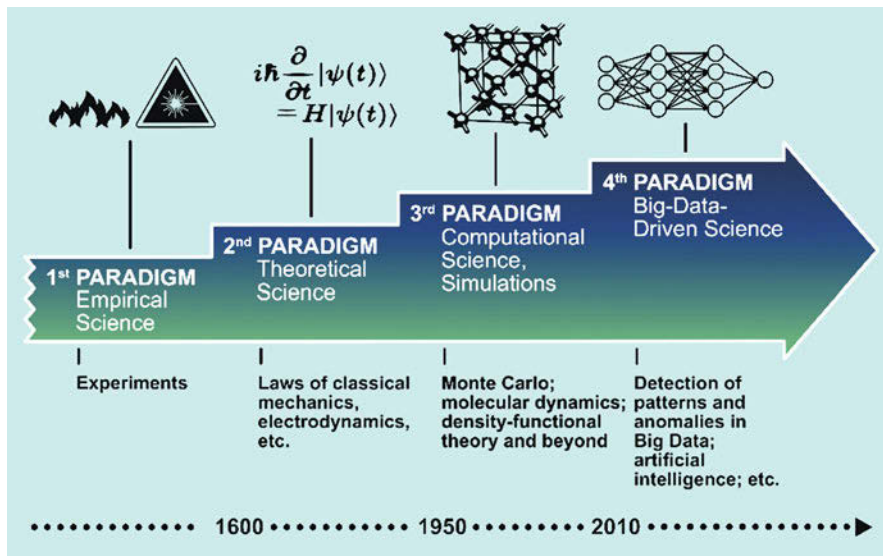
**Fig. 1** Development of research paradigms of materials science and engineering

Carlo method was introduced by Metropolis et al. (1953) and Alder and Wainwright (1958, 1962, 1970), and Rahman (1964) introduced molecular dynamics. Hohenberg and Kohn (1964) and Kohn and Sham (1965) laid the foundation of density-functional theory (DFT) (see ▶ Chap. 11, "Recent Developments in Density Functional Approximations" by Li and Burke) in the mid-1960s. All these developments enabled computer-based studies and analysis of thermodynamics and statistical mechanics on the one hand and of quantum mechanical properties of solids and liquids on the other hand. They define the beginning of computational materials science, what is nowadays considered the third paradigm. Herewith "computer experiments" were introduced, i.e., simulations, whose results are often treated and analyzed analogously to those of experimental studies. Initially developed independently, the fields of electronic-structure theory and statistical mechanics and thermodynamics are now growing together (Reuter et al. 2005 and references therein). Likewise, in parallel to DFT, many-body techniques based on Green functions were developed (Hedin 1965), which are now synergistically interleaved with DFT to form the forefront of electronic-structure theory, including excitations.

Today, big data and artificial intelligence revolutionize many areas of our life, and materials science is no exception (Gray 2007; Agrawal and Choudhary 2016). Jim Gray had probably first discussed this fourth paradigm arguing explicitly that big data reveal correlations and dependencies that cannot be seen when studying small data sets. Let us generalize the "big data" concept by noting the complexity of materials science (and others sciences as well): The number of potential but initially unknown descriptive parameters that characterize or identify the properties and functions of interest may be very big. Thus, the diversity and complexity of

mechanisms represents a big data issue in materials science as well. A further important difference to the second paradigm is that we accept that many materials properties, i.e., patterns and correlations in big data, cannot be described in terms of a closed mathematical formulation, as they are governed by several, intermingled theoretical concepts and multilevel, intricate processes. As a consequence, such patterns represent and advance knowledge, but they do not necessarily reflect understanding.

What is our vision for exploiting the fourth research paradigm of our field? Figure 2 provides a schematic view on it: The space of different chemical and structural materials is practically infinite. However, when asking, e.g., for high-performance thermal-barrier coatings, there are just a few suitable materials. Thus, in terms of functional materials, the materials space is sparsely populated as most of the already existing or in the future synthesized materials are irrelevant. Hence, the relevant data are a *statistically exceptional minority,* and if this situation is ignored, a statistical analysis that assigns the same importance to all data may well be problematic. Finding regions or patterns that correspond to materials with superior functional performance requires the identification of appropriate descriptors, noted as $d_1$ and $d_2$ in Fig. 2. Obviously, in general, the dimensionality will likely be higher than just two. At this point, Fig. 2 is just a sketch, and as for most properties, the appropriate descriptors are unknown, and the patterns remain hidden. The scientific challenge is to find appropriate descriptors. Let us emphasize, as the actually relevant data is not big enough (in most cases of interest in materials science), it is important if not crucial to use as much domain knowledge as possible. Compressed sensing, subgroup discovery, and other methods of artificial intelligence are able to identify these descriptors and patterns, and we will address these methods in Sect. 3 below.

We close this section by noting the radically decreasing time scales of new developments and paradigms: sixth millennium BCE, 1600, 1950, and 2010. Thus, a next research paradigm may be just ahead of us.
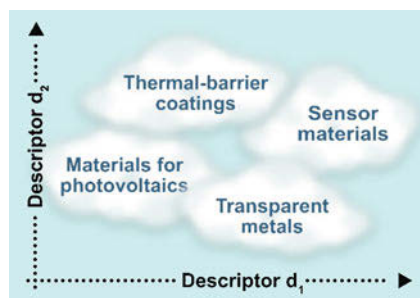


**Fig. 2** Big data contain correlations and structure that are not visible in small data sets. Finding descriptors that determine a specific property or function of a material is a crucial challenge. Once this is in place, we will be able to machine learn the data and eventually draw maps of materials. In difference to the sketch, these maps will be typically of higher dimension than just two

## 3 Extensive Data Sharing: Why and How?

Data are a crucial raw material of this century. Our community is producing materials data by CPU-intensive calculations since many years. Typically, the results are stored on PCs, workstations, or local computers, and most of these data are not used and often even thrown away, though the information content could well be significant. The field is now slowly changing its scientific culture toward *Open Science* and *Open Data*, and there are many reasons for doing so. Open access and FAIRness of data implies that data can be used by anyone, not just by the experts who develop or run advanced computer codes. If data were openly available (and well described), many more people would work with the data, e.g., computer scientists, applied mathematicians, analytic condensed matter scientists, and more. We will be surprised what people will do with data when they are made available, probably using tools that the present computational materials science community does not even know.

Let us mention one example, a "crowd sourcing" data analytics competition at the Internet platform Kaggle (Kaggle/NOMAD2018 2018). This competition addressed the search for novel transparent and semiconducting materials using a data set of $(Al_xGa_yIn_z)_2O_3$ compounds (with $x + y + z = 1$). The aim of this challenge was to identify the best machine-learning model for the prediction of two key physical properties that are relevant for optoelectronic applications: the electronic bandgap energy and the crystal formation energy. These target properties were provided for 2400 systems covering 7 different crystal space groups and unit cells ranging from 10 to 80 atoms. Six hundred additional systems made the test set for the competition.

The competition was launched in December 2017, and when it finished, 2 months later, an impressive number of 883 solutions had been submitted by researchers or research teams from around the world, employing different methods. Interestingly, the three top approaches, summarized in a recent publication (Sutton et al. 2019), adopted completely different descriptors and regression models. For example, the winning solution employed a crystal-graph representation to convert the crystal structure into features by counting the contiguous sequences of unique atomic sites of various lengths (called *n*-grams) and combined this with kernel ridge regression (KRR). This *n*-grams approach was new for materials science.

To enable the envisioned success of big data-driven materials science, the field obviously needs a suitable data infrastructure that facilitates efficient collection, data description in terms of metadata, and sharing of data. For the field of computational materials science, this was developed by the NOMAD Center of Excellence (NOMAD) which also instigated comprehensive data sharing. The synergetic relationship with other major key data bases, in particular AFLOW (see also ▶ Chap. 82, "The AFLOW Fleet for Materials Discovery" by Toher et al. in this handbook, and Curtarolo et al. (2012), Calderon et al. (2015)), Materials Project (see also Jain et al. (2013) and ▶ Chap. 81, "The Materials Project: Accelerating Materials Design Through Theory-Driven Data and Tools," by Jain et al. in this handbook), and OQMD (see also Saal et al. (2013)), made it the biggest repository in the field. Uploaded data files are tagged by a persistent identifier (PID), and users

can request a DOI (digital object identifier) to make data citable. Downloading does not require any registrations and refers to the Creative Commons License.

So what is behind FAIR? What does it mean for computational materials science?

The *F* stands for *Findable*. Making research data open and keeping them for at least 10 years are now requested by many research organizations. Seen from a practical point of view, it is also useful to avoid doubling of work and thus save human and computational resources and energy. Since individual researchers create their data on various platforms – from workstations to compute clusters to high-performance computing (HPC) centers – it is often impossible to find a student's or postdoc's data, some time after she/he has left the research group. Besides matters of organization, issues may be related to automatic erasure of data in HPC centers, missing permissions on local machines, data protection, and alike. Clearly, making data findable requires a proper data infrastructure, including documentation, metadata, search engines, and hardware. This is one of the reasons why the NOMAD Repository and its metadata were established (see Meta info at NOMAD, NOMAD 2014).

The *A* stands for *Accessible*. Accessibility in materials science has different facets. First, we should not forget about the proper hardware that allows for swift access to data. Second, we need to provide application programming interfaces (APIs). To make data fully accessible requires an important additional step, namely, the formal description of the data, i.e., its metadata that also consider the metadata interrelations. This connects to the *I* in FAIR.

The *I* stands for *Interoperable*. Here we need to consider in a first place the extreme heterogeneity of computational materials data. The wider community is using about 40 different, major computer codes (considering electronic-structure, molecular-dynamics, and quantum-chemistry packages for materials) that differ in various aspects of methodology and implementation. Consequently, the necessity arises to make their results comparable, which is a major challenge not only in the sense that they need to be brought to a common format and common units; we also recall that one quantity may be named differently in different (sub-)communities or one and the same expression may have a different meaning in one or the other area. Thus, "dictionaries" are needed to translate between them. Obviously, if one would restrict everything to just one computer program package, translations or conversion are not necessary. However, then a significant part of the available data and, even more importantly, of the community would be missed. In this sense, the NOMAD concept is general and in fact orthogonal to essentially all other data repositories that typically focus on just one computer code (see Draxl and Scheffler 2019).

Still, we need to ask if one can operate upon all available data in a meaningful way. Apart from formats and units, there are more severe restrictions that may hamper such important undertaking. These concern the computational parameters – and consequently the numerical precision – that are used in different calculations. We recall here that for extensive sharing, all open data can be considered valid in the sense that, when the code and code version are provided, output files correspond to the provided input files in a reproducible way. Nevertheless, data have been created for different purposes which may require different levels of convergence in terms
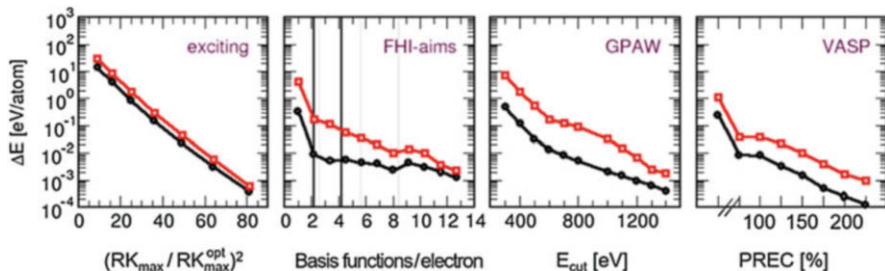
**Fig. 3** Convergence of total energies of 71 elemental solids with increasing basis-set quality for 4 different electronic-structure codes. Red symbols indicate the materials with maximal error, and black symbols display averages over all 71 materials. All values refer to the respective fully converged results. $RK_{max}$ is the LAPW-specific parameter defining the basis set size, $E_{cut}$ is the GPAW cutoff energy, and PREC is the VASP-specific parameter for basis set quality. The black vertical lines for FHI-aims indicate light (left) and tight (right) settings, and the gray lines indicate Tier1, Tier2, and Tier3 basis sets (from left to right). Note that the errors are displayed at a logarithmic scale
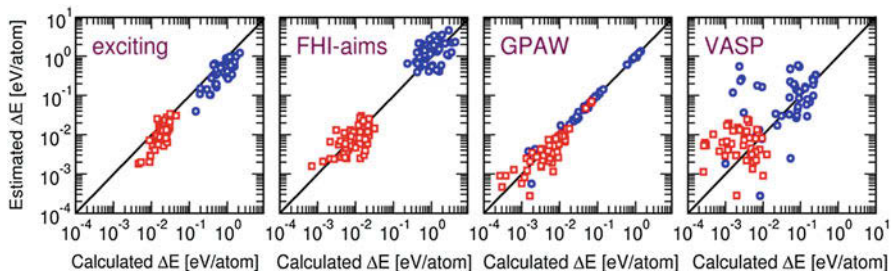


**Fig. 4** Predicted versus actual errors in total energy for 63 binary alloys, obtained from the errors in elemental solids for four different electronic structure codes. Blue data are from calculations performed at low-precision settings, and red data are from "standard quality" settings. Note that they are close to the straight black line. For details see Carbogno et al. (2020)

of basis set size and alike (e.g., the determination of the atomic geometry may need less stringent parameters than details in the electronic band structure. (see also the discussion of Figs. 3 and 4)). More than that, we may even ask whether different codes aiming at the solution of one and the same problem with "safe" settings give the same results. For the latter, we point to the community effort led by Stefaan Cottenier (Lejaeghere et al. 2016), where the equations of state for 71 elemental solids were calculated with many different *ab initio* electronic structure packages. Over a time span of a few years, it turned out that upon improvements of codes, basis sets, and in particular pseudopotentials, all codes lead to basically the same answer. In fact, in the list at https://molmod.ugent.be/deltacodesdft, one has to choose a reference where obviously an all-electron code is the natural choice, and the precision of the all-electron packages WIEN2k, exciting, and FHI-

aims are practically identical. Such investigations are extremely helpful and have set the stage toward a culture of benchmarking, which had been established in quantum chemistry for molecules already decades ago. The study by Lejaeghere et al. (2016) is, however, only the beginning. Clearly, other properties, like energy barriers, bandgaps, spectra, etc., and systems, like point defects, surfaces, interfaces, and inorganic/organic hybrid materials, etc., will be much less forgiving than total energies of simple bulk solids and will make discrepancies more obvious. Therefore, more investigations along these lines are on the way.

While the above comparison (Lejaeghere et al. 2016) could only be made with parameter sets that represent full convergence, calculations performed in daily life are often far from this optimal case and are, in fact, often sufficient. This situation obviously leads to the question how to compare and operate on calculations that have been performed with different settings, e.g., in terms of basis sets, meshes for Brillouin zone integrations, and alike. Below, it is shown that this is, in fact, possible.

Let us demonstrate that fully converged results of *complex* materials can be estimated by learning from errors of calculations of *simple* materials (Carbogno et al. 2020). Four different codes have been employed for this study: two very different all-electron codes and two projector augmented wave/plane-wave codes. These are exciting (Gulans et al. 2014) and FHI-aims (Blum et al. 2009) and GPAW (Enkovaara et al. 2010) and VASP (Kresse and Furthmüller 1996). Sources for discrepancies of different calculations in publications and/or the various data repositories are incomplete basis sets, approximate treatment of the **k**-space integration, the use of pseudopotentials, and more. Since incomplete basis sets are, indeed, the most severe issue, Fig. 3 shows how the total energies for fixed geometries change as a function of basis set quality for the 71 elemental solids adopted from Lejaeghere et al. (2016). The red symbols in Fig. 3 mark the materials exhibiting the largest error, and the black symbols refer to the average taken over all materials. The error, $\Delta E$, is defined for each material with respect to the fully converged value obtained with settings as or even more precise than the ones used in Lejaeghere et al. (2016). In all cases, the error decreases systematically from the order of 1 eV for small basis sets down to meV precision for the high-quality basis set. The fact that not all computer codes show a monotonous behavior for small basis set sizes reflects the specific characteristics of different methods and implementations (see Carbogno et al. 2020 for a detailed discussion). Based on these errors of the elemental solids, the authors estimated the errors arising from the basis set incompleteness in multicomponent systems, utilizing a simple analytic model, i.e., by linearly combining the respective errors of the constituents (elemental solids) obtained with the same settings:

$$\overline{\Delta E_{\text{tot}}} = \frac{1}{N} \sum_N N_i \Delta E_{\text{tot}}^i. \tag{1}$$

Here $N_i$ is the number of atoms of species $i$ present in the compound, and $\Delta E_{\text{tot}}^i$ is the error of this species in the respective elemental solid. This model was applied to 63 binary solids that were chosen such to cover the chemical space (1 for each element with atomic number up to 71, without noble gases). The approach is validated by comparing the estimated errors to those of corresponding DFT calculations for these binaries, as depicted in Fig. 4. The authors of this work find a clear correlation between predicted and actual errors for all four codes (with the details depending on peculiarities of the specific codes). For more in-depth analysis, also including relative energies and ternary alloys and an extensive discussion about the role of the employed method/basis set, we refer the reader to Carbogno et al. (2020). In essence, the conclusion can be drawn that even when total energies are calculated for a not converged basis set, the energetics for the fully converged case can be extrapolated. This study sets the stage for data coming from different sources fulfilling an important condition for the *I*.

The *R* stands for *Reusable*. In fact, we prefer the term *repurposable* that gives a better impression about what is meant in the materials science context. It means that we can use data that has been created for some specific scientific question, in a different connection. Indeed, one and the same material can be considered for various applications. So why should a researcher working on one aspect not allow another researcher to use the same data for focusing on a different topic? Let us illustrate this with the example of $TiO_2$ which is an important support material for heterogeneous catalysis. The detailed results are not only useful for researchers working in this area but also in a different context. For example, for photovoltaics $TiO_2$ is a component of dye-sensitized solar cells. And, as another example we note that $TiO_2$ is used as pigment in paints and in cosmetic products.

It is generally agreed that research results obtained in academia should be published. In view of what was discussed above, it should be a duty to publish all the results, i.e., making also the complete data underlying a publication available. This has been said by many people, funding agencies, politicians, and research journals. Indeed, a few research journals have started to request that all details are uploaded at a certified repository. Obviously, as mentioned above, data must be connected to established metadata and workflows.[1] However, to fully apply such concepts, the necessary data infrastructure hardly exists so far. Let us cite from a recent Nature Editorial "Empty rhetoric over data sharing slows science" (Nature editorial 2017): "Everyone agrees that there are good reasons for having open data. It speeds research, allowing others to build promptly on results. It improves replicability. It enables scientists to test whether claims in a paper truly reflect the whole data set. It helps them to find incorrect data. And it improves the attribution of credit

---

[1] In technical terms "workflow" refers to the sequence and full description of operations for creating the input file and performing the actual calculations. Important workflow frameworks that allow to automatically steer, analyze, and/or manage electronic structure theory calculation are ASE (atomic simulation environment) (Larsen et al. 2017), Fireworks (Jain et al. 2015), AFLOW (Calderon et al. 2015), and AiiDa (Pizzi et al. 2016).

to the data's originators. But who will pay? And who will host?" – and further "Governments, funders and scientific communities must move beyond lip-service and commit to data-sharing practices and platforms." For computational materials science though, NOMAD had already implemented an appropriate program (Draxl et al. 2017, Draxl and Scheffler 2018, 2019).

## 4    Artificial-Intelligence Concepts for Materials Science Data

We are using "artificial intelligence (AI)" as umbrella term for computational methods that "learn from experience". As already mentioned, for materials science, the complexity of the actuating mechanisms is big, but the number of relevant data is typically on the lower side. Zhang and Ling (2018) recently addressed this issue in terms of the degree of freedom of the model and the prediction precision. In a similar sense, we argue that a proper analysis of data needs to consider at least some aspects of the causality that drives the correlations of interest, i.e., one needs to include domain knowledge in the learning process in order to achieve a trustworthy description, interpretability, and possibly even deeper understanding of the cause behind the structure or patterns in the data. Specifically, we recall that from the practically infinite number of possible materials, only 10 or 100 may be relevant for a certain purpose. Standard machine learning tools address a description of the crowd or the majority and optimize the overall root-mean-square error or alike and also introduce a regularization to avoid overfitting and/or to achieve a smooth description. As a result, it is likely that statistically exceptional data are concealed. But only these may carry the information we are interested in. The lower the employed domain knowledge is, the larger is the amount of data that is needed for the learning process, and it may happen that data are fitted but predictions and even interpolations are not reliable. AI is a wide and interdisciplinary field, and machine learning (learning from data) and compressed sensing (originating from signal compression; primarily aiming at the identification of a low-dimensional descriptor) are important subdomains.

As noted above (see the discussion of Fig. 2), big data may reveal correlations (structure and patters) if and only if the data are arranged in a proper way, e.g., represented by appropriate descriptors. These correlations can be found by AI, but the identification of such correlations does not necessarily go along with deeper insight or understanding. To some extent we like to argue that the wish for insight and understanding is often overrated. This is well documented by the Periodic Table of the Elements that may arguably be considered as one of the most impacting achievements for chemistry, condensed matter physics, engineering, and biophysics. When Mendeleev published his table in 1871, based on knowledge of 63 atoms (their weights and chemical behavior), there was no understanding of the deeper cause behind the structure of the table (Scerri 2008; Pyykkö 2012). Still, the table predicted the existence of at that time unknown atoms. The underlying causality, i.e., that the different rows reflect the different number of nodes of the radial wave functions of the outer valence electrons and that the different columns refer to the

number of valence electrons, was unknown when the table was created. It only was understood about 50 years later, when the shell structure of electrons in atoms was described by quantum mechanics.

Thus, identifying correlations, structures, and patterns in (big) data is an important step by its own. When the relationship between a property of interest, $P$, and a set of useful descriptive parameters (the descriptors $d_1$, $d_2$, ... – also called representation) is known, graphs as in Fig. 2, or approximate equations, can be obtained for the relationship $P(d_1, d_2, ...)$. For the example of the Periodic Table, the descriptors are the row and column numbers. Obviously, as the number of possible materials is practically infinite, building a map as in Fig. 2 is a highly demanding task, of much higher complexity than building the Periodic Table of the Elements.

How to find the descriptors for materials properties? The general and complete descriptor for a quantum mechanical materials problem is given by the position of all atoms, the nuclear numbers, and the total number of electrons: $\{\boldsymbol{R}_I, Z_I\}$, $N^e$. These descriptive parameters fully identify the many-body Hamiltonian, but "learning" the properties that result from a given Hamiltonian is a very demanding goal. Thus, the amount of data needed for training (fitting) a materials property or function directly in terms of $\{\boldsymbol{R}_I, Z_I\}$, $N^e$ is typically unrealistically high. Instead, for choosing proper descriptors, that relate to the actuating mechanism of the property of interest, we distinguish three concepts: (a) the descriptor may be selected out of a huge, systematically created pool of candidates; (b) the descriptor may be built in the machine learning step in an abstract manner; and (c) one may just hand pick a known descriptor assuming that with many data the actual choice may be not so important. Concept (a) will be discussed below when we describe compressed sensing and subgroup discovery. Concept (b) is realized in neural network approaches which, in the learning step, minimizes an objective function that quantifies the difference between the predicted and the correct (known) data. Through this minimization, the weights (i.e., parameters) of the neural network are optimized (Hinton 2006; Hinton et al. 2006), and in this way, the network learns the descriptors. Doren and coworkers and Lorenz et al. (Blank et al. 1995; Lorenz et al. 2004, 2006) have shown early examples of representing potential-energy surfaces of materials by neural networks. Hellström and Behler describe recent advances in ▶ Chap. 31, "Neural Network Potentials in Materials Modeling." Concept (c) is probably the most widely used approach. Example descriptors that encode the chemical and geometrical information are Coulomb matrices (Rupp et al. 2012; Hansen et al. 2013), scattering transforms (Hirn et al. 2015), diffraction patterns (Ziletti et al. 2018), bags of bonds (Hansen et al. 2015), many-body tensor representation (Huo and Rupp (2017), smooth overlap of atomic positions (SOAP) (Bartók et al. 2010, 2013), and several symmetry-invariant transformations of atomic coordinates (Seko et al. 2017; Schütt et al. 2014; Faber et al. 2015). This concept is nicely described in ▶ Chap. 87, "Machine Learning of Atomic-Scale Properties Based on Physical Principles" by Ceriotti et al. The recent work by Xie and Grossmann (2018) may be viewed as a combination of concepts b) and c) where the input to their neural network is provided in terms of crystal graphs, an encoding of the above noted $\{\boldsymbol{R}_I, Z_I\}$ geometry and chemistry information.

A systematic understanding of the suitability of various machine learning (ML) models and thorough benchmarking studies are still lacking in materials science. A first step was only recently addressed in terms of a public data analytics competition that was hosted on the Internet platform Kaggle using a data set of 3000 $(Al_xGa_yIn_z)_2O_3$ compounds ($x + y + z = 1$), already mentioned in Sect. 3 above (see Sutton et al. 2019).

Some caution may be appropriate. In general, an observed correlation will have a causal reason – provided that it is supported by a sufficiently large data set (Pearl 2009). Thus, a correlation that is described by the function $P(d_1, d_2, \dots)$ may reflect that $(d_1, d_2, \dots) = \boldsymbol{d}$ are the actuators: $\boldsymbol{d} \to P$. However, it could well be that the reverse is true: $P \to \boldsymbol{d}$. Thirdly, it is possible that there is an "external master," $M$, who controls both $\boldsymbol{d}$ and $P$, with no direct relationship between $\boldsymbol{d}$ and $P$. And fourthly, the data may be selected with a significant bias of the researcher or research community. We fear that the latter may be happening much more frequently than realized. But then the observed correlation may just reflect this bias. All this needs to be kept in mind when tools of artificial intelligence are applied to big (or not so big) data and when we ask for interpretability or even causality.

Let us add another warning about big data of materials science. The number of possible materials is practically infinite, and we like to identify new materials that have better performance or functionality than the materials that are used today. Clearly, the amount of available data in materials science is getting big though from the few (about 250,000) inorganic materials that have been synthesized up to now, we often just know the atomic structure and hardly their electronic, elastic, or other properties. Getting more and more data, however, does not imply that all the data are relevant for all properties of interest. Materials science shows a high diversity, i.e., it is ruled by a significant number of different properties and mechanisms, and experience seems to show that at the end, the number of materials that are good for a certain group of functions is very small. For example, if we ask for a highly transparent materials with excellent heat conductivity and scratch resistance, there is probably nothing better than and nothing even close to diamond. Another example is the recent study by Singh et al. (2019) who studied 68,860 candidate materials for the photocatalytic reduction of $CO_2$. Only 52 of them turned out to be possibly relevant. In general, it is likely that in the few "high-performance materials", a mechanism is active (or inactive) that is not relevant (or dominant) in the majority of materials. Thus, the amount of available data may be big but the number of *relevant* data, i.e., data that contain the needed information, is small. In simple words, in materials science and engineering, we are often looking for "needles in a hay stack," i.e., for very few systems that are statistically exceptional, in some ways. Fitting all data (i.e., the hay) with a single, global model may average away the specialties of the minority (i.e., the needles). Thus, we need methods that identify the relevant, possibly small, statistically exceptional subgroups in the large amount of data right away.

Let us sketch this challenge for kernel methods of machine learning approaches which presently play a significant role in the field. The property of interest is written as a sum over a large, appropriate subset of all known data $j$

$$P(\boldsymbol{d}) = \sum_{j=1}^{N} c_j K(\boldsymbol{d}, \boldsymbol{d_j}).\tag{2}$$

There are many options for the kernel $K$. A most popular and very successful choice is the Gaussian kernel

$$K(\boldsymbol{d}, \boldsymbol{d_j}) = \exp\left(-\left(\boldsymbol{d} - \boldsymbol{d_j}\right)^2 / 2\sigma_j^2\right).\tag{3}$$

Fitting a set of say $N = 100,000$ known data is achieved by determining 100,000 coefficients by minimizing

$$\text{Min}\left\{\sum_{i=1}^{N}\left(\hat{P}_i - P(\boldsymbol{d_i})\right)^2 - \lambda\,||m||\right\}.\tag{4}$$

Here $\left\{\hat{P}_i\right\}$ are the actually known data that should be fitted, and we also introduced a regularization term which prevents overfitting and creates a result that does not go exactly trough the known data points but is smooth. This regularization is noted as "norm of the applied model, $m$." Figure 5 gives an example of such fitting/machine learning approach. Let us, at first, ignore that these are two types of data (noted as squares and circles): Obviously, fitting $N$ data points with a function that contains $N$ free parameters must work, but the regularization creates some uncertainty (a smooths curve), and the interpretability of the very many determined coefficients is typically lacking. Figure 5 also shows a subgroup (statistically exceptional data), together with its selector. These data are not described well by the employed kernel approach. Unfortunately, typically we don't know which data



**Fig. 5** Sketch for a machine learning method, i.e., fit of 1,000 data points (full line) and the confidence interval, which contains 69% of all the data using Gaussian process regression. Also noted is a subgroup (encircled by a blue dashed line) that is statistically exceptional but not treated correctly in the global machine-learned description. The selector equation is noted in blue (Boley 2017). For details see the subsection on subgroup discovery
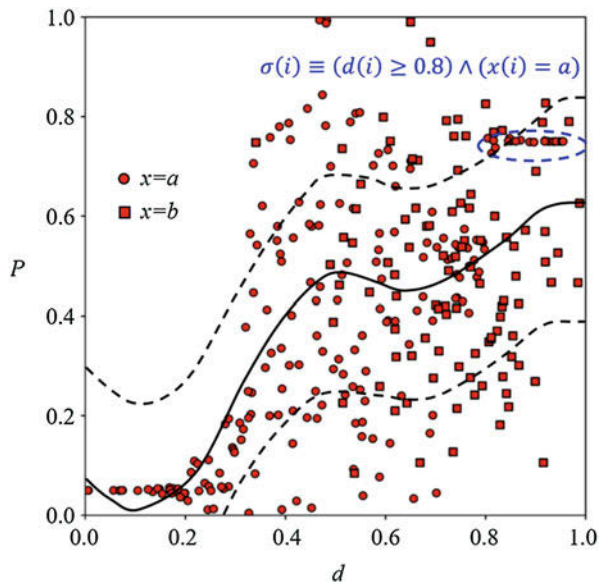
are relevant for the scientific question of interest and which are not. Otherwise, a weighted sampling could be imposed (Chawla et al. 2002 and references therein). The example of Fig. 5 reveals that the general statement "more data provide a better description" does typically not apply to ML for materials science as it may just mean, add more irrelevant information (more hay) to the information pool (the hay stack). Obviously, this will not help to find the needles. Alternatively, could we turn this around? Can we attempt to fit the hay and then consider the few materials that are distinguished by a high fitting error as an interesting subgroup that contains the needles? The difficulty here is that materials are very heterogeneous, and this heterogeneity is not just restricted to the direct hay-needle comparison. Also the "hay" is heterogeneous, and a high fitting error could also result from over- or underfitting and is not necessarily correlated with the target properties of interest.

Nevertheless, whenever we attempt a global description, machine learning is a great tool. ▶ Chapter 86, "Quantum Machine Learning in Chemistry and Materials" by Huang et al. gives an excellent description, and the abovementioned work on metal oxides (Sutton et al. 2019) is a good example.

Two interpretability-driven approaches have recently been adopted by materials science. These are *subgroup discovery* on the one hand and *compressed sensing* on the other. Let us introduce them briefly.

## 4.1    Subgroup Discovery

As noted above, a global model addressing the quantitative description of the entire data set may be difficult to interpret. For many requirements in materials research, local models that identify and describe subgroups would be advantageous. For illustration (see Fig. 6), a globally optimal regression model could predict a negative relationship between the data (Fig. 6 left). However, a subgroup discovery analysis may reveal that there are two statistically exceptional data groups (indicated by blue and red color in the right part of the figure). Thus the relationship in the data set does not have a negative slope (the global model) but positive slopes (the two subgroups). As a physical example, the transition metals of the Periodic Table are a subgroup, and the actinides, lanthanides, and halogens are other subgroups. Thus,



**Fig. 6** Left, data points and a global regression machine learning model predicting a negative relationship between material properties $y_1$ and $y_2$. Right, subgroup discovery identifies statistically exceptional regions marked as red and blue, and machine learning of these two regions exhibits positive slopes.
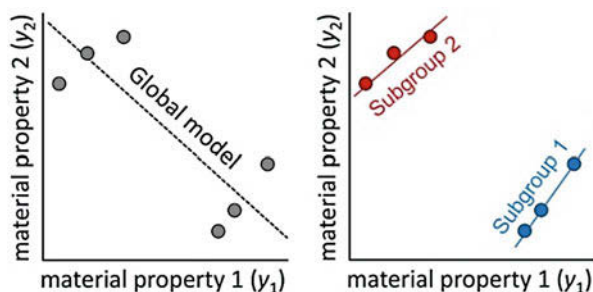
identification of subgroups is useful to gain an understanding of similarities of and differences between families of materials.

The concepts of subgroup discovery (SGD) was introduced in the early 1990s, when the advent of large databases motivated the development of explorative and descriptive analytics tools as an interpretable complement to global modeling (Duivesteijn et al. 2016; Klösgen 1996; Atzmueller 2015; Herrera et al. 2011; Siebes 1995; Wrobel 1997; Friedman and Fisher 1999). Simply speaking, the identification of subgroups is built on three components:

(i) The use of a description language for identifying subpopulations within a given pool of data. These are Boolean expressions, e.g., "the ionization potential of atom A minus the ionization potential of atom B is smaller than $X$" where $X$ is a number that may be chosen iteratively. These expressions are called selectors.

(ii) The definition of utility functions that formalize the interestingness (quality) of a subpopulation. This may include requests as "the bandgap of the material is between 1.1 and 1.5 eV" and "the cohesive energy is larger than 3 eV."

(iii) The design of search algorithms to find selectors that describe the subpopulations of interest (e.g., Goldsmith et al. 2017).

Figure 7 illustrates the approach for a recent study of heterogeneous catalysis: finding potential catalysts that can transform the greenhouse gas $CO_2$ into useful chemicals or fuels (Mazheika et al. 2019). This study concentrated on metal oxides and realized that a global model (fitting all the data at once) failed to provide an accurate description. However, searching for subgroups by considering several potential indicators for a good catalytic activity and many potential selectors reveals several subgroups that are statistically exceptional.
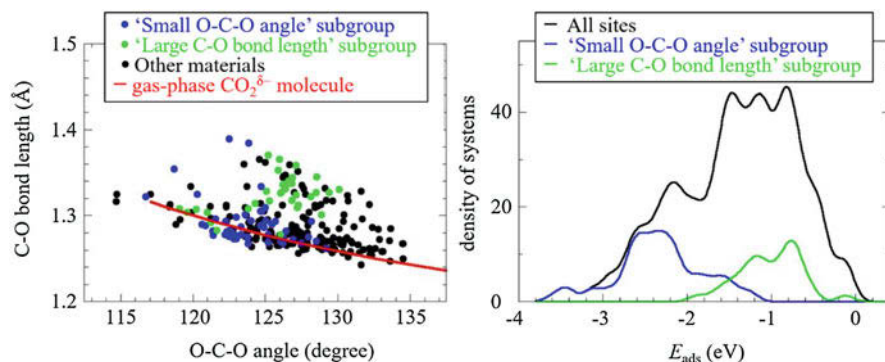


**Fig. 7** Two subgroups of pristine oxide materials in a data set that describes the $CO_2$ adsorption. The blue subgroup is characterized by "small" angles of the O-C-O molecule (the neutral free molecule is linear (180°)). And the green subgroup is characterized by at least one large C-O bond length (the bond lengths of the neutral free molecules are 1.17 Å). Both subgroups reflect a weakening of the bond, but only the green one correlates with a good catalytic activity. Right: density of systems of the full data set and of the two subgroups as function of adsorption energy

Only one of them (marked in green in Fig. 7) contains the known catalytically active materials. Details can be found in Mazheika et al. (2019). The study identified a new indicator for the catalytic $CO_2$ activation, and it provided several suggestions for new potentially good catalysts.

## 4.2    Compressed Sensing and the SISSO Approach

As noted in the discussion of Fig. 2, finding a descriptor (two-dimensional in Fig. 2), i.e., the set of parameters that capture the underlying mechanism of a given materials property or function, is the key, intelligent step toward identification of structure or patterns in (big) data. This central role of the descriptor was only recently addressed explicitly and systematically in the works of Ghiringhelli and coworkers (Ghiringhelli et al. 2015, 2017b; Ouyang et al. 2018). These authors recast the descriptor identification challenge into a systematic formulation using compressed sensing (CS).

The CS approach had been originally introduced for efficiently reproducing a high-quality signal from a very small set of observations (Candès and Wakin 2008; Nelson et al. 2013; Candès et al. 2006; Candro et al. 2006; Donoho 2006). Mathematically, this can be sketched as follows. Given a set of data $P_1, P_2, \ldots P_N$, where $i = 1-N$ labels different materials (or different conditions), CS finds the sparse solution $\boldsymbol{c}$ of an underdetermined system of linear equations

$$P(\boldsymbol{d}_i) = \sum_{k=1}^{M} \hat{c}_k \, d_{ki} \; . \tag{5}$$

$\{d_{ki}\} = \boldsymbol{D}$ is called the sensing matrix with the number of rows $k = 1-M$ significantly bigger than the number of columns, $i = 1-N$. Thus, the sensing matrix is built from $N$ vectors (the columns), each of length $M$. Material $i$ is characterized by vector $i$, i.e., by $k = 1-M$ values (descriptive parameters), $d_{ki}$. Equation (5) corresponds to Eq. (2) when the linear kernel is used. If most elements of the vector $\hat{\boldsymbol{c}}$ were zero, specifically when the number of nonzero elements of $\hat{\boldsymbol{c}}$ is smaller than $N$, the dimensionality of the problem is reduced (Candès et al. 2006; Donoho 2006; Candès and Wakin 2008). In order to achieve this reduction, the coefficients $\hat{c}_k$ are determined by solving Eq. (4) with the norm $||m||$ taken as the $l_0$ norm of $\hat{\boldsymbol{c}}$. The norm zero of a vector is defined as the number of nonzero elements. Thus, the regularization $\lambda||\hat{\boldsymbol{c}}||_0$ is a constraint that favors solutions for $\hat{\boldsymbol{c}}$ where most elements of $\hat{\boldsymbol{c}}$ are zero. However, using the norm zero poses a combinatorically extensive problem, and it has been shown that this is (asymptotically) NP hard. As a consequence it has been suggested to approximate the norm zero by the norm $l_1$, and a popular approach to it is LASSO (least absolute shrinkage and selection operator) (Tibshirani 1996). For materials science this has been introduced by Ghiringhelli and coworkers (Ghiringhelli et al. 2015, 2017b).

Thus, the idea behind the compressed sensing approach is to offer many descriptor candidates and then let the optimization approach (Eq. (4)) find out which of these candidates are relevant. Since Eq. (5) is linear, it is necessary that the offered descriptor candidates contain the potential nonlinearities. Consequently, different descriptor candidates, i.e., different columns of the sensing matrix, may become correlated. Furthermore, the dimension of the sensing matrix increases rapidly with the number of data points, and as LASSO requires that the matrix is stored, the approach is getting intractable. These problems have been recently tackled by Ouyang and coworkers (Ouyang et al. 2018) by solving the $l_0$ challenge in an iterative approach called SISSO (sure independence screening and sparsifying operator). Interestingly, the mentioned correlations are not causing problems, and the number of candidate descriptors can be increased in SISSO to many billions and even trillions. Initially, from the previously mentioned "basic descriptors" $\{\boldsymbol{R}_I, Z_I\}$, $N^e$, only $Z_I$-derived quantities were used explicitly, e.g., the ionization potentials of the atoms, the electron affinities, and information about the extension of the atomic wave functions. Then, a combination of algebraic/functional operations is recursively performed for extending the space of descriptor candidates. Details are described in Ouyang et al. (2018). Clearly, when different structures are considered or different charge states $\{\boldsymbol{R}_I\}$, $N^e$-related features are needed as well.

## 5 Outlook

Computational materials science took off with impressive early work by Moruzzi et al. (1978) on various properties of metals and by Cohen and coworkers (Yin and Cohen 1982) on the cohesion and phase transition of silicon and germanium (see ▶ Chap. 3, "Modeling Solids and Its Impact on Science and Technology" by M. Cohen in this handbook, in particular Fig. 4). A number of computer codes for solving the Kohn-Sham equations have been developed since then, initially involving approximations like pseudopotentials (removing the core electrons, creating smooth potentials) or introducing touching or slightly overlapping atom-centered spheres in which the potential was sphericalized. During the 1980s significant advancements in the original pseudopotential approach have been made (see the work of Vanderbilt and coworkers: Garrity et al. 2014 and references therein), and all-electron codes have been developed that don't require a shape approximation for the potentials (e.g., Wimmer et al. 1981; Blaha et al. 1990; Gulans et al. 2014; Blum et al. 2009). The work by Lejaeghere et al. (2016) provides a nice overview of the precision of modern electronic-structure codes for elemental bulk solids, also demonstrating how to involve the community. Clearly, this kind of work is important and needs to be extended to more complicated structures and compositions, defects, surfaces, and interfaces. Work in this direction is underway, as are studies for advanced electronic structure methods, like the *GW* approach (van Setten et al. 2015). Furthermore, the field urgently needs benchmarks for the various numerical approximations and for exchange-correlation potentials in order to address also accuracy, not only numerical precision. The MSE (materials science and engineering) project (Zhang et al. 2019)

is a promising step in this direction. Without all this, data-driven science will be limited in its capabilities.

Computational materials science is presently dominated by the third research paradigm (cf. Fig. 1), but advancements in AI methods has been significant in recent years, and the fourth paradigm is playing an increasing role. Still, at present there is more hype than realism in what AI can do. Much of this relates to the fact that the amount of relevant or independent data is often not really big. Machine learning techniques can help a lot when general trends are of interest and when one needs to fit and predict "the behavior of a big crowd" (see, e.g., the methods used in the Kaggle competition for predicting properties of transparent conductors Sutton et al. 2019). Often, the sensible needs of materials science and engineering are, however, different: We are typically not looking for a crowd behavior, but we are searching for materials with extraordinary performance on certain functions or properties, often even a combination of several properties. There are typically just a few, statistically exceptional, suitable materials among the enormous number of possible materials (already existing ones and those that will be synthesized in the future). However, in many cases we don't know how to identify this "class" of potentially interesting functional materials. How can we distinguish which data/materials are relevant and which are not? Learning about less than 0.01% relevant materials from thousands or millions of irrelevant data is obviously problematic, and standard methods, which optimize the regularized root-mean-square error, even emphasize the importance of the irrelevant data, while surpassing the special cases. If the data could be grouped in "a majority classes" and a "minority class," methods have been developed to deal with the problem (Chawla et al. 2002 and references therein). However, often these classes are unknown, and advancements of the subgroup discovery concept for the materials science domain are urgently needed.

What is missing at present? Let us list some issues:

- Close coupling of materials property prediction with stability analysis and prediction of routes reproducible synthesis.
- High-throughput studies of *metastable* materials and of the lifetimes of these metastable states.
- Materials under real conditions ($T$, $p$, and reactive environment): stability and properties. This very much concerns multiscale modeling with robust, error-controlled links with knowledge of uncertainty between the various simulation methodologies. This has been often stated in the past but is still not fully realized.
- Error estimates of calculations in terms of numerical approximations (basis sets, pseudopotentials, etc.) for specific properties (structure, elastic and electronic properties, etc.).
- Computations beyond standard DFT as, for example, coupled-cluster methods for calculations for solids (possibly also getting prepared for quantum computers).
- Complete description of scientific results accounting for the heterogeneity of data, i.e., to improve and complement present metadata definitions. While significant progress has been made for computational data from the many computer

codes and the development of corresponding metadata (Ghringhelli et al. 2016, 2017a; Meta Info at NOMAD), it is now urgent that the same is being achieved also for experimental data. The latter challenge is even bigger than the first. The sample material used in the experimental study corresponds to the input file of a calculation; the experimental condition ($T$, $p$, environment) and the experimental equipment to the computer code. A not fully solved challenge is the complete characterization of the sample material. In fact, this will often be impossible. Obviously, closely coupled to the definition of metadata is the description of workflows in the sample preparation and running of the experiment.

The field is just developing the methods for the fourth paradigm. The learning curve connecting paradigms 1, 2, 3, and 4 is apparently getting steep. Thus the next paradigm may be close, even though the fourth has not been developed well, so far. What could be the next paradigm? Considering that "the future is already here – it's just not very evenly distributed" (Gibson 1999), it may be hintingly visible already today. We guess that it may be virtual reality with direct and instantaneous connection to new calculations or a merger of theory (predictions and analysis) and experiment. There are exciting times ahead of us.

# References

AFLOW, Automatic FLOW for materials discovery, http://aflowlib.org/

Agrawal A, Choudhary A (2016) Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science. APL Mater 4:053208

Alder BJ, Wainwright TE (1958) Molecular dynamics by electronic computers. In: Prigogine I (ed) International symposium on transport processes in statistical mechanics. Wiley, New York, pp 97–131

Alder BJ, Wainwright TE (1962) Phase transition in elastic disks. Phys Rev 127:359–361

Alder BJ, Wainwright TE (1970) Decay of velocity autocorrelation function. Phys Rev A 1:18–21

Atzmueller M (2015) Subgroup discovery. WIREs Data Min Knowl Discov 5:35

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403

Bartók AP, Kondor R, Csányi G (2013) On representing chemical environments. Phys Rev B 87:184115

Blaha P, Schwarz K, Sorantin P, Trickey SB (1990) Full-potential, linearized augmented plane wave programs for crystalline systems. Comp Phys Commun 59:399

Blank TB, Brown SD, Calhoun AW, Doren DJ (1995) Neural network models of potential energy surfaces. J Chem Phys 103:4129

Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M (2009) Ab initio molecular simulations with numeric atom-centered orbitals. Comput Phys Commun 180: 2175–2196

Boley M (2017) Private communications. In the figure, the Gaussian radial basis function (rbf) kernel was used plus a 0.1 noise component: k(a,b)=rbf(a,b | scale=0.2) + 0.1 delta(a,b)

Calderon CE, Plata JJ, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl MJ, Hart G, Nardelli MB, Curtarolo S (2015) The AFLOW standard for high-throughput materials science calculations. Comput Mater Sci 108:233

Candès EJ, Wakin MB (2008) An introduction to compressive sampling. IEEE Signal Proc Mag 25:21

Candès EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52:489

Candro EJ, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory 52:489

Carbogno C, Thygesen KS, Bieniek B, Draxl C, Ghiringhelli LM, Gulans A, Hofmann OT, Jacobsen KW, Lubeck S, Mortensen JJ, Strange M, Wruss E, Scheffler M (2020) Numerical quality control for DFT-based materials databases. Preprint to be published

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321

Curtarolo S, Setyawan W, Hart GLW, Jahnatek M, Chepulskii RV, Taylor RH, Wanga S, Xue J, Yang K, Levy O, Mehl MJ, Stokes HT, Demchenko DO, Morgan D (2012) AFLOW: an automatic framework for high-throughput materials discovery. Comput Mater Sci 58:218

Donoho DL (2006) Compressed sensing. IEEETrans InformTheory 52:1289

Draxl C, Scheffler M (2018) NOMAD: the FAIR concept for big-data-driven materials science. MRS Bull 43:676

Draxl C, Scheffler M (2019) The NOMAD laboratory: from data sharing to artificial intelligence. J Phys Mater 2:036001

Draxl C, Illas F, Scheffler M (2017) Open data settled in materials theory. Nature 548:523

Duivesteijn W, Feelders AJ, Knobbe A (2016) Exceptional model mining: supervised descriptive local pattern mining with complex target concepts. Data Min Knowl Discov 30:47

Enkovaara J, Rostgaard MJJ, Chen J, Dułak M, Ferrighi L, Gavnholt J, Glinsvad C, Haikola V, Hansen HA, Kristoffersen HH, Kuisma M, Larsen AH, Lehtovaara L, Ljungberg M, Lopez-Acevedo O, Moses PG, Ojanen J, Olsen T, Petzold V, Romero NA, Stausholm-Møller J, Strange M, Tritsaris GA, Vanin M, Walter M, Hammer B, Häkkinen H, Madsen GKH, Nieminen RM, Nørskov JK, Puska M, Rantala TT, Schiøtz J, Thygesen KS, Jacobsen KW (2010) Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. J Phys Condens Matter 22:253202

Faber F, Lindmaa A, von Lilienfeld OA, Armiento R (2015) Crystal structure representations for machine learning models of formation energies. Int J Quantum Chem 115:1094

Friedman JH, Fisher NI (1999) Bump hunting in high-dimensional data. Statistics and Computing 9:123

Garrity KF, Bennett JW, Rabe KM, Vanderbilt D (2014) Pseudopotentials for high-throughput DFT calculations. Comput Mater Sci 81:446–452

Ghiringhelli LM, Vybiral J, Levchenko SV, Draxl C, Scheffler M (2015) Big data of material science: critical role of the descriptor. Phys Rev Lett 114:105503

Ghiringhelli LM, Carbogno C, Levchenko S, Mohamed F, Huhs G, Lüder M, Oliveira M, Scheffler M (2016) Towards a common format for computational materials science data. Psi-k Scientific Highlight of the Month No. 131. http://psi-k.net/download/highlights/Highlight_131.pdf

Ghiringhelli LM, Carbogno C, Levchenko S, Mohamed F, Hus G, Lüder M, Oliveira M, Scheffler M (2017a) Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. npj Comput Mater 3:46

Ghiringhelli LM, Vybiral J, Ahmetcik E, Ouyang R, Levchenko SV, Draxl C, Scheffler M (2017b) Learning physical descriptors for material science by compressed sensing. New J Phys 19:023017

Gibson WF (1999) "The Science in Science Fiction" on Talk of the Nation (30 Nov 1999, Timecode 11:55). Available via NPR. https://www.npr.org/2018/10/22/1067220/the-science-in-science-fiction or https://www.npr.org/programs/talk-of-the-nation/1999/11/30/12966633/

Goldsmith BR, Boley M, Vreeken J, Scheffler M, Ghiringhelli LM (2017) Uncovering structure-property relationships of materials by subgroup discovery. New J Phys 19:013031

Gray J (2007) The concept of a fourth paradigm was probably first discussed by J. Gray at a workshop on January 11, 2007 before he went missing at the Pacific on January 28, 2007. See: Hey T, Tansley S, Tolle K (eds) (2009) The fourth paradigm, data intensive discovery. Microsoft Research, Redmond, Washington 2009, ISBN 978–0–9825442-0-4

Gulans A, Kontur S, Meisenbichler C, Nabok D, Pavone P, Rigamonti S, Sagmeister S, Werner U, Draxl C (2014) Exciting: a full-potential all-electron package implementing density-functional theory and many-body perturbation theory. J Phys Condens Matter 26:363202

Hansen K, Montavon G, Biegler F, Fazli S, Rupp M, Scheffler M, von Lilienfeld OA, Tkatchenko A, Müller K-K (2013) Assessment and validation of machine learning methods for predicting molecular atomization energies. J Chem Theory Comput 9:3404

Hansen K, Biegler F, Ramakrishnan R, Pronobis W, von Lilienfeld OA, Müller K-R, Tkatchenko A (2015) Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. J Phys Chem Lett 6:2326

Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139:A796

Herrera F, Carmona CJ, González P, del Jesus MJ (2011) An overview on subgroup discovery: foundations and applications. Knowl Inf Syst 29:495

Hinton GE (2006) Reducing the dimensionality of data with neural networks. Science 313: 504–507

Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18:1527

Hirn M, Poilvert N, Mallat S (2015) Quantum energy regression using scattering transforms. https://arxiv.org/abs/1502.02077

Hohenberg P, Kohn W (1964) Inhomogeneous Electron Gas. Phys Rev 136:B864

Huo H, Rupp M (2017) Unified representation for machine learning of molecules and crystals. https://arxiv.org/abs/1704.06439

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) The materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:011002

Jain A, Ong SP, Chen W, Medasani B, Qu X, Kocher M, Brafman M, Petretto G, Rignanese GM, Hautier G, Gunter D, Persson KA (2015) FireWorks: a dynamic workflow system designed for high-throughput applications. Concurr Comput: Pract Exper 27:5037–5059

Kaggle/Nomad2018 (2018) Predicting transparent conductors – predict the key properties of novel transparent semiconductors https://www.kaggle.com/c/nomad2018-predict-transparent-conductors

Klösgen W (1996) Explora: a multipattern and multistrategy discovery assistant. In: Advanced techniques in knowledge discovery and data mining. American Association for Artificial Intelligence, Menlo Park, p 249

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138

Kresse G, Furthmüller J (1996) Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169

Larsen AH, Mortensen JJ, Blomqvist J, Castelli IE, Christensen R, Dułak M, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Jensen PB, Kermode J, Kitchin JR, Kolsbjerg

EL, Kubal J, Kaasbjerg K, Lysgaard S, Maronsson JB, Maxson T, Olsen T, Pastewka L, Peterson A, Rostgaard C, Schiøtz J, Schütt O, Strange M, Thygesen KS, Vegge T, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW (2017) The atomic simulation environment – a Python library for working with atoms. J Phys Condens Mat 29:273002

Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google flu: traps in big data analysis. Science 343:1203

Lejaeghere K, Bihlmayer G, Björkamn T, Blaha P, Blügel S, Blum V, Caliste D, Castelli IE, Clark SJ, Corso AD, de Gironcoli S, Deutsch T, Dewhurst JK, Di Marco I, Draxl C, Dulak M, Eriksson O, Flores-Livas JA, Garrity KF, Genovese L, Giannozzi P, Giantomassi M, Goedecker S, Gonze X, Grånäs O, Gross EKU, Gulans A, Gygi F, Hamann DR, Hasnip PJ, Holzwarth NAW, Iuşan D, Jochym DB, Jollet F, Jones D, Kresse G, Koepernik K, Küçükbenli E, Kvashnin YO, Locht ILM, Lubeck S, Marsman M, Marzari N, Nitzsche U, Nordström L, Ozaki T, Paulatto L, Pickard CJ, Poelmans W, Probert MIJ, Refson K, Richter M, Rignanese G-M, Saha S, Scheffler M, Schlipf M, Schwarz K, Sharma S, Tavazza F, Thunström P, Tkatchenko A, Torrent M, Vanderbildt D, van Setten MJ, Speyvroeck VV, Wills JM, Yates JR, Zhang G-X, Cottenier S (2016) Reproducibility in density functional theory calculations of solids. Science 351: aad3000

Lorenz S, Groß A, Scheffler M (2004) Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. Chem Phys Lett 395:210

Lorenz S, Scheffler M, Groß A (2006) Descriptions of surface chemical reactions using a neural network representation of the potential-energy surface. Phys Rev B 73:115431

Materials Project. https://materialsproject.org

Mazheika A, Wang Y, Ghiringhelli LM, Illas F, Levchenko SV, Scheffler M (2019) *Ab initio* data analytics study of carbon-dioxide activation on semiconductor oxide surfaces. http://arxiv.org/abs/1912.06515

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087

Moruzzi VL, Janak JF, Williains AR (1978) Calculated electronic properties of metals. Pergamon, New York

Nature editorial (2017) Not-so-open data. Nature 546:327. Empty rhetoric over data sharing slows science https://www.nature.com/news/empty-rhetoric-over-data-sharing-slows-science-1.22133

Nelson IJ, Hart GLW, Zhou F, Ozolins V (2013) Compressive sensing as a paradigm for building physics models. Phys Rev B 87:035125

NOMAD (2014) The concept of the NOMAD Repository and Archive (NOMAD) was developed in 2014 (see e.g. the discussion in Ghiringhelli et al. 2016), independently and parallel to the "FAIR Guiding Principles" (Wilkinson et al. 2016). Interestingly, the essence is practically identical. However, the accessibility of data in NOMAD goes further than meant in the FAIR Guiding Principles, as for searching and even downloading data from NOMAD, users don't even need to register

NOMAD, The NOMAD (Novel Materials Discovery) Center of Excellence (CoE) was launched in November 2015. https://nomad-coe.eu, https://youtu.be/yawM2ThVlGw

OQMD, Open quantum materials database. http://oqmd.org/

Ouyang R, Curtarolo S, Ahmetcik E, Scheffler M, Ghiringhelli LM (2018) SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. Phys Rev Mat 2:083802

Pearl J (2009) Causality: models, reasoning and inference, 2nd edn. Cambridge University Press, New York

Pizzi J, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci 111:218–230

Pyykkö P (2012) The physics behind chemistry and the periodic table. Chem Rev 112: 371–384

Rahman A (1964) Correlations in the motion of atoms in liquid argon. Phys Rev 136:A405–A411

Reuter K, Stampfl C, Scheffler M (2005) *Ab Initio* atomistic thermodynamics and statistical mechanics of surface properties and functions. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht, pp 149–194

Rupp M, Tkatchenko A, Müller K-R, von Lilienfeld OA (2012) Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 108:058301

Saal J, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). JOM 65:1501

Scerri ER (2008) The periodic table: its story and its significance. Oxford University Press, New York. ISBN 978-0-19-530573-9

Schütt KT, Glawe H, Brockherde F, Sanna A, Müller K-R, Gross EKU (2014) How to represent crystal structures for machine learning: towards fast prediction of electronic properties. Phys Rev B 89:205118

Seko A, Hayashi H, Nakayama K, Takahashi A, Tanaka I (2017) Representation of compounds for machine-learning prediction of physical properties. Phys Rev B 95:144110

Siebes A (1995) Data surveying foundations of an inductive query language. KDD-95 proceedings. AAAI Press, Montreal, p 269

Singh AK, Montoya JH, Gregoire JM, Persson KA (2019) Robust and synthesizable photocatalysts for $CO_2$ reduction: a data-driven materials discovery. Nat Commun 10:443

Slater JC (1937) Wave functions in a periodic potential. Phys Rev 51:846

Slater JC (1953) An augmented plane wave method for the periodic potential problem. Phys Rev 92:603

Slater JC (1965) Quantum theory of molecules and solids, Symmetry and energy bands in crystals, vol 2. McGraw-Hill, New York

Slater JC (1967) Quantum theory of molecules and solids, insulators, semiconductors and metals, vol 3. McGraw-Hill, New York

Slater JC, Johnson KH (1972) Self-consistent-field Xα cluster method for polyatomic molecules and solids. Phys Rev B 5:844

Sutton C, Ghiringhelli LM, Yamamoto T, Lysogorskiy Y, Blumenthal L, Hammerschmidt T, Golebiowski J, Liu X, Ziletti A, Scheffler M (2019) Crowd-sourcing materials-science challenges with the NOMAD 2018 Kaggle competition. npj Comput Mater 5:1–11. https://doi.org/10.1038/s41524-019-0239-3

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B 58:267

van Setten MJ, Caruso F, Sharifzadeh S, Ren X, Scheffler M, Liu F, Lischner J, Lin L, Deslippe JR, Louie SG, Yang C, Weigend F, Neaton JB, Evers F, Rinke P (2015) *GW*100: benchmarking $G_0W_0$ for molecular systems. J Chem Theory Comput 11:5665

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Monsal B (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018

Wimmer E, Krakauer H, Weinert M, Freeman AJ (1981) Full-potential self-consistent linearized-augmented-plane-wave method for calculating the electronic structure of molecules and surfaces: $O_2$ molecule. Phys Rev B 24:864

Wrobel S (1997) An algorithm for multi-relational discovery of subgroups. In: Komorowski J, Zytkow J (eds) Principles of data mining and knowledge discovery: first European symposium, PKDD'97, Trondheim, Norway, 24–27 June 1997. Springer, Berlin, p 78

Xie T, Grossman JC (2018) Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. Phys Rev Lett 120:145301

Yin MT, Cohen ML (1982) Theory of static structural properties, crystal stability, and phase transformations: application to Si and Ge. Phys Rev B 26:5668

Zhang IY, Logsdail AJ, Ren X, Levchenko SV, Ghiringhelli L, Scheffler M (2019) Test set for materials science and engineering with user-friendly graphic tools for error analysis: systematic benchmark of the numerical and intrinsic errors in state-of-the-art electronic-structure approximations. New J Phys 1:013025

Zhang Y, Ling C (2018) A strategy to apply machine learning to small datasets in materials science. npj Comput Mater 4:25

Ziletti A, Kumar D, Scheffler M, Ghiringhelli LM (2018) Insightful classification of crystal structures using deep learning. Nat Commun 9:2775

# TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter

**6**

## Federica Agostini, Basile F. E. Curchod, Rodolphe Vuilleumier, Ivano Tavernelli, and Eberhard K. U. Gross

## Contents

F. Agostini
Laboratoire de Chimie Physique, University Paris-Saclay, Orsay, France
e-mail: federica.agostini@u-psud.fr

B. F. E. Curchod
Department of Chemistry, Durham University, Durham, UK
e-mail: basile.f.curchod@durham.ac.uk

R. Vuilleumier
PASTEUR, Département de chimie, École normale supérieure, PSL University, Sorbonne
Université, CNRS, Paris, France
e-mail: rodolphe.vuilleumier@ens.fr

I. Tavernelli
Zurich Research Laboratory, IBM Research GmbH, Rüschlikon, Switzerland
e-mail: ita@zurich.ibm.com

E. K. U. Gross (✉)
Max Planck Institute of Microstructure Physics, Halle, Germany
e-mail: hardy@mpi-halle.mpg.de

**Abstract**

Time-dependent density functional theory (TDDFT) is currently the most efficient approach allowing to describe electronic dynamics in complex systems, from isolated molecules to the condensed phase. TDDFT has been employed to investigate an extremely wide range of time-dependent phenomena, as spin dynamics in solids, charge and energy transport in nanoscale devices, and photoinduced exciton transfer in molecular aggregates. It is therefore nearly impossible to give a general account of all developments and applications of TDDFT in material science, as well as in physics and chemistry. A large variety of aspects are covered throughout these volumes. In the present chapter, we will limit our presentation to the description of TDDFT developments and applications in the field of quantum molecular dynamics simulations in combination with trajectory-based approaches for the study of nonadiabatic excited-state phenomena. We will present different quantum-classical strategies used to describe the coupled dynamics of electrons and nuclei underlying nonadiabatic processes. In addition, we will give an account of the most recent applications with the aim of illustrating the nature of the problems that can be addressed with the help of these approaches. The potential, as well as the limitations, of the presented methods is discussed, along with possible avenues for future developments in TDDFT and nonadiabatic dynamics.

## 1    Introduction

Photoinduced isomerization processes, photosynthetic and photovoltaic energy conversion phenomena, and charge and energy transport through molecular junctions are all typical examples of, so-called, nonadiabatic processes. Nonadiabatic processes are characterized by a strong coupling between electronic and nuclear motion; in fact, nuclear motion is responsible for inducing electronic (nonadiabatic) transitions, and in turn, the time evolution of the electronic states also affects the nuclear dynamics at very short timescales (down to a few tens of fs). In this nonadiabatic regime, thus when the Born-Oppenheimer approximation breaks down, performing (quantum) molecular dynamics simulations is tremendously challenging. Accurate electronic structure properties are required to describe electronic dynamics and to correctly *drive* the nuclear evolution. Identifying regions of nuclear configuration space where the electronic states are coupled, as avoided crossings and conical intersections, is essential to predict quantum yields. Efficient evolution techniques have to be employed to describe nuclear motion in order, for instance, to determine final molecular structures or to account for possible quantum effects. Therefore, theoretical and numerical developments need to address the problem from the perspective of both electronic structure theory and nuclear quantum dynamics.

Perhaps the most celebrated method to investigate excited electronic states is time-dependent density functional theory (TDDFT). TDDFT offers an in principle exact formalism for propagating the time-dependent electronic density and, within

linear-response theory, for calculating excitation energies as well as critical excited-state properties. It is therefore without any surprise that TDDFT became the electronic structure method of choice to be coupled with nonadiabatic dynamics. Particularly successful has been the combination of TDDFT, employed to solve the electronic problem, with the description of nuclear motion in terms of trajectories that evolve or hop between coupled (electronic) potential energy surfaces. The most well-known method is Tully's "fewest switches" trajectory surface hopping (Tully 1990), which has evolved into a widely used and successful technique. The mean-field Ehrenfest dynamics is often employed to investigate explicitly the electronic dynamics, combined, for example, with real-time TDDFT (Tavernelli et al. 2005; Tavernelli 2006). Full multiple spawning (Martínez et al. 1996; Martínez and Levine 1997; Ben-Nun and Martínez 1998, 2002; Ben-Nun et al. 2000; Hack et al. 2001; Virshup et al. 2008) propagates coupled Gaussian functions along classical trajectories, whereas the coupled-trajectory mixed quantum-classical (CT-MQC) scheme (Min et al. 2015) derived from the exact factorization (Abedi et al. 2010) is based on the propagation of trajectories along a time-dependent potential energy surface (Abedi et al. 2013a). Other techniques like the quantum-classical Liouville equation (Kapral and Ciccotti 1999; Nielsen et al. 2000; Kapral 2006), Bohmian dynamics (Wyatt et al. 2001; Lopreore and Wyatt 2002; Rassolov and Garashchuk 2005; Curchod and Tavernelli 2013), variational multiconfiguration Gaussians (Worth et al. 2004, 2008; Lasorne et al. 2006, 2007; Mendive-Tapia et al. 2012; Richings et al. 2015), multiconfigurational Ehrenfest (Shalashilin 2010; Saita and Shalashilin 2012; Makhov et al. 2017), or linearization approaches to compute time-correlation functions (Bonella and Coker 2005; Huo and Coker 2012; Dunkel et al. 2008) have been also proposed for nonadiabatic dynamics. Despite their differences, all the methods mentioned above are rooted in the Born-Huang representation of the total molecular wavefunction, i.e., an expansion in an infinite sum over the correlated Born-Oppenheimer electronic states. In contrast, the recently introduced exact factorization of the time-dependent molecular wavefunction offers a paradigm shift in our perception of nonadiabatic dynamics, away from the Born-Huang picture, and blazes a trail for the development of nonadiabatic techniques away from Born-Oppenheimer concepts. It is important to mention here that ensembles of trajectories, when properly constructed via the method of characteristics (Agostini et al. 2018), can represent, in principle arbitrarily closely, the solution of the underlying partial differential equation. Practical implementations, however, involve further-going approximations where, for instance, interference and tunneling effects are neglected or only approximately taken into account. The advantage of trajectory-based method is that they circumvent the enormous numerical effort associated with quantum wave packets propagation techniques, such as multiconfiguration time-dependent Hartree approach (MCTDH) (Meyer et al. 1990; Burghardt et al. 1999; Wang and Thoss 2003; Meyer and Worth 2003). By its very nature, this approach requires the computation of the relevant potential energy surfaces (PESs) and corresponding couplings before the actual propagation of nuclear wave packets. This clearly implies an important computational effort that limits the applicability of this method to a small number of degrees of freedom (up to ∼10). In addition,

the determination of the relevant degrees of freedom to include in the dynamics can also become a challenging problem, which requires some a priori knowledge of the "active" vibrational modes involved in the dynamics. Such wavefunction-based nonadiabatic approaches are beyond the scope of this chapter and will not be discussed further.

The goal of this chapter is to present in a self-contained manner the key theoretical concepts and equations of the most important methods cited above, starting from the electronic structure problem and (LR-)TDDFT up to nuclear dynamics methods like surface hopping, Ehrenfest dynamics, and ab initio multiple spawning. To contrast with these Born-Huang-based methods, we also present the formalism of the exact factorization and introduce the reader to the first mixed quantum-classical algorithm derived from this formalism, coined coupled-trajectory mixed quantum-classical (CT-MQC) dynamics.

## 2  Coupled Electron-Nuclear Dynamics in Molecules

In molecules and condensed phase systems, the time evolution of interacting electrons and nuclei is described by the time-dependent Schrödinger equation:

$$\hat{H}\Psi(\mathbf{r}, \mathbf{R}, t) = i\hbar\partial_t\Psi(\mathbf{r}, \mathbf{R}, t), \tag{1}$$

where the electron-nuclear wavefunction $\Psi(\mathbf{r}, \mathbf{R}, t)$ describes the state of the system over time and $\hat{H}$ is the molecular Hamiltonian, i.e.:

$$\hat{H}(\mathbf{r}, \mathbf{R}) = \sum_{v=1}^{N_n} \frac{-\hbar^2}{2M_v}\nabla_v^2 + \hat{T}_e(\mathbf{r}) + V_{ee}(\mathbf{r}) + V_{nn}(\mathbf{R}) + V_{en}(\mathbf{r}, \mathbf{R})$$

$$= \sum_{v=1}^{N_n} \frac{-\hbar^2}{2M_v}\nabla_v^2 + \hat{H}_{BO}(\mathbf{r}, \mathbf{R}). \tag{2}$$

Here, $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_{N_{el}})$, $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_{N_n})$, $N_{el}$ is the number of electrons and $N_n$ the number of nuclei. The first term on the right-hand side of Eq. (2) is the nuclear kinetic energy, with $\nabla_v$ indicating the spatial derivative with respect to the position of the nucleus $v$ and $M_v$ its mass, whereas $\hat{H}_{BO}$ is the so-called Born-Oppenheimer (BO), or electronic, Hamiltonian. $\hat{H}_{BO}$ is defined as the sum of the electronic kinetic energy, $\hat{T}_e$; the electron-electron, $\hat{V}_{ee}$; the nucleus-nucleus $V_{nn}$; and the electron-nucleus, $V_{en}$, interactions.

Usually, the problem is reformulated adopting the Born-Huang expansion of the molecular wavefunction in the adiabatic basis. The adiabatic, or BO, states, $\varphi_{\mathbf{R}}^{(k)}(\mathbf{r})$, are defined as the eigenfunctions of the BO Hamiltonian:

$$\hat{H}_{BO}(\mathbf{r}, \mathbf{R})\varphi_{\mathbf{R}}^{(k)}(\mathbf{r}) = \varepsilon_{BO}^{(k)}(\mathbf{R})\varphi_{\mathbf{R}}^{(k)}(\mathbf{r}), \tag{3}$$

with eigenvalues $\varepsilon_{BO}^{(k)}(\mathbf{R})$. The electronic time-independent problem is diagonalized at each nuclear position $\mathbf{R}$; thus the eigenfunctions and eigenvalues depend on $\mathbf{R}$. Nuclear positions are interpreted here as parameters, which label both the electronic states and the electronic energies. If Eq. (3) is solved for all nuclear configurations, $\varepsilon_{BO}^{(k)}(\mathbf{R})$ identify the so-called BO potential energy surfaces (PESs). In the Born-Huang expansion of the electron-nuclear wavefunction,

$$\Psi(\mathbf{r}, \mathbf{R}, t) = \sum_k \chi_k(\mathbf{R}, t) \varphi_{\mathbf{R}}^{(k)}(\mathbf{r}), \qquad (4)$$

the coefficients $\chi_k(\mathbf{R}, t)$ clearly depend on nuclear positions and on time. These coefficients can be interpreted as the nuclear contributions corresponding to the electronic states included in the sum and can be also referred to as nuclear wave packets. In fact, it can be easily proven that the nuclear density, defined as the integral of $|\Psi(\mathbf{r}, \mathbf{R}, t)|^2$ over electronic coordinates,

$$\int d\mathbf{r} \, |\Psi(\mathbf{r}, \mathbf{R}, t)|^2 = \sum_k |\chi_k(\mathbf{R}, t)|^2, \qquad (5)$$

can be written as the sum of adiabatic contributions, $|\chi_k(\mathbf{R}, t)|^2$. Here, the orthogonality of the BO states

$$\int d\mathbf{r} \varphi_{\mathbf{R}}^{(l)*}(\mathbf{r}) \varphi_{\mathbf{R}}^{(k)}(\mathbf{r}) = \left\langle \varphi_{\mathbf{R}}^{(l)} \middle| \varphi_{\mathbf{R}}^{(k)} \right\rangle_{\mathbf{r}} = \delta_{lk} \qquad (6)$$

has been used.

The Born-Huang expansion (4) is inserted in Eq. (1), which is then projected on $\varphi_{\mathbf{R}}^{(k)*}(\mathbf{r})$ and integrated over $\mathbf{r}$. A set of partial differential equations are derived for the expansion coefficients:

$$\left[ \sum_v^{N_n} \frac{-\hbar^2}{2M_v} \nabla_v^2 + \varepsilon_{BO}^{(k)}(\mathbf{R}) \right] \chi_k(\mathbf{R}, t) + \sum_l \mathscr{F}_{kl}(\mathbf{R}) \chi_l(\mathbf{R}, t) = i\hbar \partial_t \chi_k(\mathbf{R}, t). \qquad (7)$$

The last term on the right-hand side is responsible for coupling the evolution of the $k$-th coefficient to all other coefficients, via the nonadiabatic couplings

$$\mathscr{F}_{kl}(\mathbf{R}) = \int d\mathbf{r} \, \varphi_{\mathbf{R}}^{(k)*}(\mathbf{r}) \left[ \sum_v^{N_n} \frac{-\hbar^2}{2M_v} \nabla_v^2 \right] \varphi_{\mathbf{R}}^{(l)}(\mathbf{r})$$

$$+ \sum_v^{N_n} \frac{1}{M_v} \left\{ \int d\mathbf{r} \, \varphi_{\mathbf{R}}^{(k)*}(\mathbf{r}) \left[ -i\hbar \nabla_v \varphi_{\mathbf{R}}^{(l)}(\mathbf{r}) \right] \right\} \cdot [-i\hbar \nabla_v], \qquad (8)$$

arising from the effect of the nuclear kinetic energy operator on the parametric dependence of the BO states on $\mathbf{R}$. In the most general case, the non-diagonal

elements of $\mathscr{F}_{kl}(\mathbf{R})$ are non-zero and induce a coupling between different electronic states due to the motion of the nuclei. The nonadiabatic coupling term is responsible for exchanging "nuclear contributions" between the electronic adiabatic states $k$ and $l$. The BO framework presented so far is widely adopted by a large community of physicists and chemists to interpret the coupled electron-nuclear problem under nonadiabatic conditions. However, such framework is not the only one, as we will discuss below.

An alternative perspective on the coupled electron-nuclear problem has been recently proposed, the exact factorization of the electron-nuclear wavefunction (Abedi et al. 2010, 2012). In this framework, we make an Ansatz different from the Born-Huang representation of the molecular wavefunction, namely:

$$\Psi(\mathbf{r}, \mathbf{R}, t) = \Phi_{\mathbf{R}}(\mathbf{r}, t)\chi(\mathbf{R}, t). \tag{9}$$

Here $\chi(\mathbf{R}, t)$ is the nuclear wavefunction and $\Phi_{\mathbf{R}}(\mathbf{r}, t)$ is the electronic wavefunction which parametrically depends on the nuclear positions and satisfies the partial normalization condition (PNC):

$$\int d\mathbf{r}\, |\Phi_{\mathbf{R}}(\mathbf{r}, t)|^2 = 1 \quad \forall\, \mathbf{R}, t. \tag{10}$$

The theorems introduced in Abedi et al. (2010, 2012) prove the existence and uniqueness of Eq. (9), up to within a $(\mathbf{R}, t)$-dependent gauge transformation. The PNC guarantees the interpretation of $|\chi(\mathbf{R}, t)|^2$ as the probability of finding the nuclear configuration $\mathbf{R}$ at time $t$ and of $|\Phi_{\mathbf{R}}(\mathbf{r}, t)|^2$ itself as the conditional probability of finding the electronic configuration $\mathbf{r}$ at time $t$ given the nuclear configuration $\mathbf{R}$.

The stationary variations (Frenkel 1934) of the quantum-mechanical action with respect to $\Phi_{\mathbf{R}}(\mathbf{r}, t)$ and $\chi(\mathbf{R}, t)$ lead to the derivation of the following equations of motion:

$$\left(\hat{H}_{BO}(\mathbf{r}, \mathbf{R}) + \hat{U}_{en}^{coup}[\Phi_{\mathbf{R}}, \chi] - \varepsilon(\mathbf{R}, t)\right)\Phi_{\mathbf{R}}(\mathbf{r}, t) = i\hbar\partial_t \Phi_{\mathbf{R}}(\mathbf{r}, t) \tag{11}$$

$$\left[\sum_{v=1}^{N_n} \frac{[-i\hbar\nabla_v + \mathbf{A}_v(\mathbf{R}, t)]^2}{2M_v} + \varepsilon(\mathbf{R}, t)\right]\chi(\mathbf{R}, t) = i\hbar\partial_t \chi(\mathbf{R}, t), \tag{12}$$

where the PNC is enforced by means of Lagrange multipliers (Alonso et al. 2013; Abedi et al. 2013b). The electron-nuclear coupling operator (Agostini et al. 2015b),

$$\hat{U}_{en}^{coup}[\Phi_{\mathbf{R}}, \chi] = \sum_{v=1}^{N_n} \frac{1}{M_v}\left[\frac{[-i\hbar\nabla_v - \mathbf{A}_v(\mathbf{R}, t)]^2}{2}\right.$$
$$\left. + \left(\frac{-i\hbar\nabla_v\chi}{\chi} + \mathbf{A}_v(\mathbf{R}, t)\right)\left(-i\hbar\nabla_v - \mathbf{A}_v(\mathbf{R}, t)\right)\right], \tag{13}$$

the time-dependent potential energy surface (TDPES) (Abedi et al. 2013a; Agostini et al. 2013; Suzuki et al. 2015; Agostini et al. 2015a; Curchod et al. 2016a; Suzuki and Watanabe 2016),

$$\varepsilon(\mathbf{R}, t) = \langle \Phi_{\mathbf{R}}(t) | \hat{H}_{BO} + \hat{U}_{en}^{coup} - i\hbar \partial_t | \Phi_{\mathbf{R}}(t) \rangle_{\mathbf{r}} , \tag{14}$$

and the time-dependent vector potential (Curchod and Agostini 2017),

$$\mathbf{A}_v(\mathbf{R}, t) = \langle \Phi_{\mathbf{R}}(t) | - i\hbar \nabla_v \; \Phi_{\mathbf{R}}(t) \rangle_{\mathbf{r}} \tag{15}$$

are responsible for the coupling between electrons and nuclei in a formally exact way. It is worth noting that the electron-nuclear coupling operator, $\hat{U}_{en}^{coup}[\Phi_{\mathbf{R}}, \chi]$, in the electronic equation (11), depends on the nuclear wavefunction and acts on the parametric dependence of $\Phi_{\mathbf{R}}(\mathbf{r}, t)$ as a differential operator. This "pseudo-operator" includes the coupling to the nuclear subsystem beyond the parametric dependence in the BO Hamiltonian $\hat{H}_{BO}(\mathbf{r}, \mathbf{R})$. The symbol $\langle \cdot \rangle_{\mathbf{r}}$ indicates an integration over electronic coordinates only. The nuclear equation (12) has the particularly appealing form of a Schrödinger equation that contains a time-dependent vector potential (15) and a time-dependent scalar potential (14) that govern the nuclear dynamics and yield the nuclear wavefunction. $\chi(\mathbf{R}, t)$ is interpreted as the nuclear wavefunction since it leads to an $N$-body nuclear density, and an $N$-body current-density, which reproduce the true nuclear $N$-body density and current-density (Abedi et al. 2012) obtained from the full wavefunction $\Psi(\mathbf{r}, \mathbf{R}, t)$.

In order to connect the Born-Huang representation to the exact factorization, the electronic wavefunction $\Phi_{\mathbf{R}}(\mathbf{R}, t)$ is expanded in terms of the BO states, similarly to what is done for the molecular wavefunction of Eq. (4), namely:

$$\Phi_{\mathbf{R}}(\mathbf{r}, t) = \sum_k C_k(\mathbf{R}, t) \varphi_{\mathbf{R}}^{(k)}(\mathbf{r}). \tag{16}$$

The expansion coefficients in Eqs. (4) and (16) are related,

$$\chi_k(\mathbf{R}, t) = C_k(\mathbf{R}, t) \chi(\mathbf{R}, t), \tag{17}$$

by virtue of the factorization (9). Additionally, the PNC can be rewritten as

$$\sum_k |C_k(\mathbf{R}, t)|^2 = 1 \quad \forall \; \mathbf{R}, t. \tag{18}$$

We point out that even in the case where the nuclear wave packet splits into more than one BO PESs, the full wavefunction is still a single product: the nuclear wave function has contributions (projections) on different BO PESs, while the electronic wavefunction is a linear combination of the adiabatic states, but still we may write:

$$\Psi(\mathbf{r}, \mathbf{R}, t) = \left( e^{\frac{i}{\hbar} S(\mathbf{R},t)} \sqrt{\sum_l |\chi_l(\mathbf{R}, t)|^2} \right) \left( \sum_k C_k(\mathbf{R}, t) \varphi_{\mathbf{R}}^{(k)}(\mathbf{r}) \right) \qquad (19)$$

where the first term in parenthesis is $\chi(\mathbf{R}, t)$, with a phase $S(\mathbf{R}, t)$ determined by fixing the gauge freedom, and the second term in parenthesis is $\Phi_{\mathbf{R}}(\mathbf{r}, t)$, using Eq. (16).

In the absence of nonadiabatic couplings in Eq. (7), the evolution equations for the coefficients $\chi_k(\mathbf{R}, t)$ decouple, and each nuclear contribution now evolves *adiabatically* according to the TDSE

$$\left[ \sum_v^{N_n} \frac{-\hbar^2}{2M_v} \nabla_v^2 + \varepsilon_{BO}^{(k)}(\mathbf{R}) \right] \chi_k(\mathbf{R}, t) = i\hbar \partial_t \chi_k(\mathbf{R}, t), \qquad (20)$$

under the effect of a potential produced only by the electrons in the adiabatic state $k$. This is the essence of the BO approximation. Analogously, in the limit of infinite nuclear masses, Eqs. (11) and (12) (Scherrer et al. 2015, 2017; Schild et al. 2016; Eich and Agostini 2016) reduce to the fundamental equations of the BO approximation, namely, the static electronic equation (3) and a nuclear evolution equation identical to Eq. (20) with $\chi_k(\mathbf{R}, t)$ replaced by the nuclear wavefunction $\chi(\mathbf{R}, t)$ of the exact factorization.

If the nonadiabatic couplings cannot be neglected, the fully coupled electron-nuclear problem, summarized in Eq. (7) or Eqs. (11) and (12), has to be solved.

Electronic dynamics is simulated at a quantum-mechanical level employing quantum chemistry approaches, either based on the electronic wavefunction or on the electronic density. If either the adiabatic or the diabatic basis is used to characterize the electronic subsystem, electronic dynamics is implied in the time evolution of the expansion coefficients (see, e.g., Eqs. (4) or (16)), since the basis functions are time-independent. On the other hand, (real-time) TDDFT yields an explicit evolution of the electronic subsystem, as the electrons are represented via their time-dependent one-body density. As we will show below, real-time TDDFT can be combined with a mean-field solution of the coupled electron-nuclear dynamics. The linear response formulation of TDDFT, instead, is able to provide information about the time-independent electronic properties, such as adiabatic forces and nonadiabatic couplings, needed for approaches based on the Born-Huang expansion. Possible extensions of TDDFT to solve the electronic equation of the exact factorization are currently under investigation (Requist and Gross 2016). Section 3 is devoted to a thorough review of the basis of TDDFT and of LR-TDDFT

Nuclear dynamics can be treated exactly or approximated at different levels, depending on the complexity of the system of interest. Simulation methods that retain the quantum character of nuclear dynamics are indeed very expensive, as the numerical cost for solving the quantum-mechanical problem scales exponentially with the number of degrees of freedom. Therefore, different strategies have been proposed over the years to make the problem numerically tractable. Quantum wave

packet propagation techniques aim at solving Eq. (7) either on grids (Lauvergnat and Nauts 2010, 2014; Sadri et al. 2012) or by expanding the nuclear wave packets $\chi_k(\mathbf{R}, t)$ on a basis where calculations are computationally cheaper (Meyer et al. 1990; Burghardt et al. 1999; Wang and Thoss 2003; Meyer and Worth 2003; Sadri et al. 2014). The major bottleneck of these approaches is the "pre-calculation" of the electronic properties, i.e., BO PESs and of the nonadiabatic couplings, needed to solve the nuclear equations. Attempts at solving exactly the coupled equations at the basis of the exact factorization are currently under investigations. On-the-fly calculations of electronic properties are instead possible, if only local nuclear information is necessary to solve (in an approximate way) Eq. (7). Full and ab initio multiple spawning methods (Martínez et al. 1996; Martínez and Levine 1997; Ben-Nun and Martínez 1998, 2002; Ben-Nun et al. 2000; Hack et al. 2001; Virshup et al. 2008), similarly to direct dynamics techniques (Worth et al. 2004, 2008; Lasorne et al. 2006, 2007; Mendive-Tapia et al. 2012; Richings et al. 2015), employ a representation of the nuclear wave packets in terms of moving Gaussian functions, which evolve along trajectories determined either variationally or classically. Trajectory-based quantum-classical schemes adopt a representation of nuclear dynamics in terms of purely classical trajectories, as in the Ehrenfest and surface-hopping methods. They are indeed numerically cheaper than the methods above, but the price to pay is sometimes the neglect of important quantum-mechanical features both in the nuclear dynamics and in the coupling between electronic and nuclear motion. Similarly to direct dynamics and full multiple spawning, evolving the nuclei along trajectories enables us to exploit the locality of classical dynamics for on-the-fly simulations, where electronic information is needed, and thus computed, only for the visited nuclear configurations. Trajectory-based solutions of Eqs. (11) and (12) have been proposed (Agostini et al. 2014; Abedi et al. 2014), and the most recent developments (Min et al. 2017) will be reviewed in Sect. 4, along with Ehrenfest dynamics (Tully 1998), trajectory surface hopping (Tully 1990; Doltsinis and Marx 2002; Böckmann et al. 2010; Jasper et al. 2004, 2006; Subotnik et al. 2013; Curchod and Tavernelli 2013; Jaeger et al. 2012; Fang and Hammes-Schiffer 1999; Tapavicza et al. 2007; Craig et al. 2005; Akimov and Prezhdo 2014), and full/ab initio multiple spawning (Martínez et al. 1996; Martínez and Levine 1997; Ben-Nun and Martínez 1998, 2002; Ben-Nun et al. 2000; Hack et al. 2001; Virshup et al. 2008).

## 3 Electronic Dynamics: Time-Dependent Density Functional Theory

### 3.1 Time-Dependent Density Functional Theory

The Hohenberg-Kohn (HK) theorem (Hohenberg and Kohn 1964) of ground-state DFT states that knowledge of the ground-state density uniquely determines the external potential of the system (up to within a trivial constant) and thus the entire electronic Hamiltonian and the associated total ground-state energy. It is important to realize that ground-state DFT in nearly all applications is intimately tied to the

BO approximation: the electronic density one calculates is the one produced by *clamped* nuclei. Then, by varying the positions of the clamped nuclei, ground-state DFT provides an efficient approach to map out the lowest BO PES and to calculate physical observables associated with the lowest BO PES, such as vibrational spectra, cohesive energies, barrier heights, etc. Higher BO PESs and the time evolution of systems strongly driven by external fields are not accessible with ground-state DFT.

In their seminal paper, Runge and Gross (1984) proved a theorem that established a 1-1 correspondence between the time-dependent density and the time-dependent external potential for systems evolving from a given initial many-electron state, $\Phi_0$. The time evolution of the many-electron wavefunction is governed by the time-dependent Schrödinger equation:

$$\hat{H}_{el}(t)\Phi(\mathbf{r}, t) = i\hbar \frac{\partial}{\partial t}\Phi(\mathbf{r}, t) \tag{21}$$

$$\Phi(\mathbf{r}, t_0) = \Phi_0(\mathbf{r})$$

with Hamiltonian

$$\hat{H}_{el}(t) = \hat{T}_e(\mathbf{r}) + V_{ee}(\mathbf{r}) + v_{ext}(\mathbf{r}, t). \tag{22}$$

The general time-dependent external potential appearing in (22) covers different scenarios: one important case is the (short-time) electron dynamics with clamped nuclei, driven by an applied laser field:

$$v_{ext}(\mathbf{r}, t) = V_{nn}(\mathbf{R}) + V_{en}(\mathbf{r}, \mathbf{R}) + \delta v_{app}(\mathbf{r}, t). \tag{23}$$

Another case is the time-dependent electric potential produced by classically propagated point-like nuclei:

$$v_{ext}(\mathbf{r}, t) = V_{nn}(\mathbf{R}(t)) + V_{en}(\mathbf{r}, \mathbf{R}(t)). \tag{24}$$

In complete detail, the Runge-Gross theorem ensures that the densities $\rho(\mathbf{r}, t)$ and $\rho'(\mathbf{r}, t)$ evolving from a common initial many-body state $\Phi_0 = \Phi(t_0)$ under the influence of two potentials $V_{ext}(\mathbf{r}, t)$ and $V'_{ext}(\mathbf{r}, t)$ will become different infinitesimally later than $t_0$ if the potentials are Taylor expandable around the initial time $t_0$ and differ by more than a purely time-dependent constant $V_{ext}(\mathbf{r}, t) \neq V'_{ext}(\mathbf{r}, t) + C(t)$. This implies that the potentials-to-densities map can be inverted:

$$\rho(\mathbf{r}, t) \rightarrow v_{ext}[\rho](\mathbf{r}, t). \tag{25}$$

The Runge-Gross proof does not depend on the particular form of the particle-particle interaction. The proof is valid for essentially any interaction, in particular also for no interaction. This establishes the map for noninteracting particles:

$$\rho(\mathbf{r}, t) \rightarrow v_s[\rho](\mathbf{r}, t), \tag{26}$$

implying that the potential $v_s(\mathbf{r}, t)$, which reproduces the interacting density, $\rho(\mathbf{r}, t)$, in a noninteracting system is uniquely defined. From now on, this unique potential $v_s[\rho](\mathbf{r}, t)$ will be called the *time-dependent Kohn-Sham potential*. The corresponding system of single-particle time-dependent Schrödinger equations

$$i\hbar \frac{\partial}{\partial t} \phi_k(\mathbf{r}, t) = \left(-\frac{1}{2}\nabla^2 + v_s(\mathbf{r}, t)\right) \phi_k(\mathbf{r}, t), \quad k = 1, \ldots, N, \tag{27}$$

whose orbitals reproduce the interacting density via

$$\rho(\mathbf{r}, t) = \sum_{i=1}^{N} |\phi_k(\mathbf{r}, t)|^2 \tag{28}$$

is called *time-dependent Kohn-Sham (TDKS) equations*.

The Runge-Gross theorem guarantees uniqueness of the potentials $v_{\text{ext}}[\rho, \Phi_0](\mathbf{r}, t)$ and $v_s[\rho, \{\phi_k^{(0)}(\mathbf{r})\}]$ for given initial many-body state $\Phi_0$ and given initial orbitals $\{\phi_k^{(0)}(\mathbf{r})\}$, respectively (Gross and Kohn 1990). Apart from uniqueness, whether or not, for a given function $\rho(\mathbf{r}, t)$, the potentials $v_{\text{ext}}(\mathbf{r}, t)$ and $v_s(\mathbf{r}, t)$ actually exist, is a separate question, known as (interacting and noninteracting) v-representability problem. This problem has been solved – once and for all – by van Leeuwen (1999), who demonstrated under mild conditions to be satisfied by the densities $\rho(\mathbf{r}, t)$ that the potentials $v_{\text{ext}}(\mathbf{r}, t)$ and $v_s(\mathbf{r}, t)$ can be constructed explicitly as solutions of the Sturm-Liouville problem. Since this is a constructive proof, the solution of the TDDFT v-representability problem is much more satisfactory than the status of the v-representability problem in ground-state DFT where a complete characterization of the domains of the $v_{\text{ext}}^{gs}(\mathbf{r}, t)$ and $v_s^{gs}(\mathbf{r}, t)$ is still lacking.

The TDKS potential in Eq. (27) is usually written in the following form:

$$v_s(\mathbf{r}, t) = v_0(\mathbf{r}, t) + v_H[\rho](\mathbf{r}, t) + v_{xc}[\rho](\mathbf{r}, t) \tag{29}$$

where $v_0(\mathbf{r}, t)$ is the given external potential of the system at hand, $v_H[\rho](\mathbf{r}, t)$ is the time-dependent Hartree potential,

$$v_H[\rho](\mathbf{r}, t) = \int d\mathbf{r}' \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} \tag{30}$$

and $v_{xc}[\rho](\mathbf{r}, t)$ is the universal exchange-correlation ($xc$) functional of TDDFT

$$v_{xc}[\rho](\mathbf{r}, t) := v_s[\rho](\mathbf{r}, t) - v_{\text{ext}}[\rho](\mathbf{r}, t) - v_H[\rho](\mathbf{r}, t). \tag{31}$$

The $xc$ functional is well-defined through the right-hand side of Eq. (31): uniqueness of $v_s[\rho]$ and $v_{ext}[\rho]$ is guaranteed by the Runge-Gross theorem, and the existence over a well-characterized domain is covered by the van Leeuwen theorem. Formally, in addition to dependence on the density $\rho(\mathbf{r}, t)$, the $xc$ potential also depends on the initial many-body state $\Phi_0$ and on the initial orbitals $\{\phi_k^{(0)}(\mathbf{r})\}$. If the initial state is a ground state, both $\Phi_0$ and $\{\phi_k^{(0)}(\mathbf{r})\}$ are functionals of the initial ground-state density $\rho_0^{gs}(\mathbf{r}, t)$ via the HK theorem, and then the time-dependent $xc$ potential becomes a functional of the time-dependent density alone. The density dependence of the exact time-dependent $xc$ functional $v_{xc}[\rho(\mathbf{r}', t')](\mathbf{r}, t)$ is nonlocal both in space and in time, i.e., the potential $v_{xc}(\mathbf{r}, t)$ at point $\mathbf{r}$ and time $t$ depends on the density values at all points $\mathbf{r}'$ and at all previous times $t' \leq t$.

An important aspect of the ground-state DFT is the HK variational principle which ensures that the total energy as functional of the density is minimized by the true ground-state density of the system at hand, and the value of the functional at the minimum is the true ground-state energy. The HK variational principle is important in two respects: first of all, the total energy is a quantity of prime interest and the variational principle guarantees that the lowest possible value is achieved. Of equal importance is the fact that the variational principle usually implies numerical stability of the iterative algorithms, such as the Kohn-Sham (KS) self-consistency cycle, because they ultimately go "downhill" in the total energy functional.

In the time-dependent case, variational principles play a less important role. First of all, the usual Frenkel variational principle of quantum mechanics

$$\delta \int_{t_0}^{t_1} dt \ \langle \Phi(t) | i\hbar \frac{\partial}{\partial t} - \hat{H} | \Phi(t) \rangle = 0 \tag{32}$$

normally does not have a minimum at the solution of the time-dependent Schrödinger equation. There is only a stationary point, and consequently there is no guarantee of the stability of the associated time propagation algorithms. Moreover, unlike the ground-state energy, the value of functional (32) in the stationary point is zero and of no physical significance. Nevertheless, a TDDFT variational principle might still be desirable for some purposes, e.g., for the optimization of constrained densities.

Straightforward combination of the Runge-Gross map with the Frenkel variational principle (32) leads to a variational formulation of TDDFT (Runge and Gross 1984) which was later found to give rise to serious inconsistencies (Gross et al. 1994). In particular, a noncausal $xc$ kernel is found. This so-called causality paradox arises from the fact that arbitrary density variations lead to variations of the wavefunction at the upper boundary $t_1$ of the Frenkel integral (32). If the variations of the wavefunction are explicitly included, the causality paradox disappears (Vignale 2008). Another way of getting rid of the upper boundary $t_1$ of the Frenkel integral (32) is to formulate the TDDFT variational principle on the Keldysh contour which maps the final time back to the initial time. This formulation of the TDDFT variational principle was achieved by van Leeuwen (1998).

## 3.2   Linear-Response TDDFT

Many applications of TDDFT deal with weak probes of the ground state of a static potential $v_0(\mathbf{r})$, mediated by a small time-dependent perturbation $\delta v_{\text{app}}(\mathbf{r}, t)$. The goal of linear-response TDDFT is to calculate the induced first-order change $\delta\rho(\mathbf{r}, t)$ in the density (Gross et al. 1996). To this end, we look at the density $\rho[v_{ext}](\mathbf{r}, t)$ as functional of the external potential and perform a functional Taylor expansion at the unperturbed ground-state potential $v_0(\mathbf{r}, t)$:

$$\rho[v_{ext}](\mathbf{r}, t) = \rho \left[ v_0 + \delta v_{app} \right] (\mathbf{r}, t) \tag{33}$$

$$= \rho[v_0](\mathbf{r}) + \int d\mathbf{r}' \int dt' \left. \frac{\delta\rho(\mathbf{r}, t)}{\delta v_{ext}(\mathbf{r}', t')} \right|_{v_0} \delta v_{app}(\mathbf{r}', t') + \dots . \tag{34}$$

The functional derivative on the right-hand side of Eq. (34), which connects the change in the density with the perturbation is of enormous physical significance. It is known as *density-density response function* and will henceforth be denoted by $\chi(\mathbf{r}, t, \mathbf{r}', t')$:

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \left. \frac{\delta\rho(\mathbf{r}, t)}{\delta v_{ext}(\mathbf{r}', t')} \right|_{v_0} . \tag{35}$$

The associated change in the density is known as *linear density response*:

$$\delta\rho(\mathbf{r}, t) = \int d\mathbf{r}' \int dt' \, \chi(\mathbf{r}, t, \mathbf{r}', t') \delta v_{app}(\mathbf{r}', t'). \tag{36}$$

Since $\chi(\mathbf{r}, t, \mathbf{r}', t')$ only depends on $t - t'$, Eq. (36) is usually Fourier-transformed to the frequency domain:

$$\delta\rho(\mathbf{r}, \omega) = \int d\mathbf{r}' \, \chi(\mathbf{r}, \mathbf{r}', \omega) \delta v_{app}(\mathbf{r}', \omega) \tag{37}$$

where, for simplicity, we use the same symbol for a function and for its Fourier transform. The poles of $\chi(\mathbf{r}, \mathbf{r}', \omega)$ provide the charge-neutral excitation energies of the unperturbed many-body system.

One may also look at the density $\rho_s[v_s](\mathbf{r}, t)$ of noninteracting particles and their density-density response function:

$$\chi_s(\mathbf{r}, t, \mathbf{r}', t') = \left. \frac{\delta\rho_s(\mathbf{r}, t)}{\delta v_s(\mathbf{r}', t')} \right|_{v_{s,0}} . \tag{38}$$

While the full interacting density-density response function (35) is very hard to evaluate (in many-body language, it is the reducible polarization propagator of the

interacting system), the noninteracting counterpart is relatively easy to calculate: Its Fourier transform reads

$$\chi_s(\mathbf{r}, \mathbf{r}', \omega) = \sum_{ij\sigma,kl\tau} \phi_{i\sigma}(\mathbf{r})\phi_{j\sigma}^*(\mathbf{r})\phi_{k\tau}(\mathbf{r}')\phi_{l\tau}^*(\mathbf{r}')\chi_{ij\sigma,kl\tau}^{s}(\omega). \tag{39}$$

with

$$\chi_{ij\sigma,kl\tau}^{s}(\omega) = \delta_{\sigma,\tau}\delta_{i,k}\delta_{j,l}\frac{f_{j\sigma} - f_{i\sigma}}{\omega - (\varepsilon_{i\sigma} - \varepsilon_{j\sigma})}, \tag{40}$$

where $\varepsilon_{i\sigma}$ are the ground-state KS orbital energies and $f_{i\sigma}$ their occupations in the ground state. Multiplying this equation – in the operator sense – from the left with $\chi_s$ and from the right with $\chi$, and performing a Fourier transform to frequency space, one obtains the following Dyson-like equation for the response function (Petersilka et al. 1996):

$$\chi(\omega) = \chi^s(\omega) + \chi^s(\omega) * f_{Hxc}(\omega) * \chi(\omega). \tag{41}$$

This equation constitutes the cornerstone of linear-response TDDFT.

Acting with the operator in Eq. (41) on an arbitrary perturbation $\delta v_{\mathrm{app}}(\mathbf{r}, \omega)$ and using the definition (37) of the linear density response, one ends up with an integral equation for the desired density response:

$$\delta\rho(\omega) = \chi_s(\omega) * \big(\delta v_{\mathrm{app}}(\omega) + f_{Hxc}(\omega) * \delta\rho(\omega)\big). \tag{42}$$

An iterative numerical solution of this equation yields the full linear density response as function of $\omega$ and was first achieved by Zangwill and Soven (1980) for atoms in the frequency regime above the continuum threshold. If one is interested in the discrete spectrum of the system, i.e., the discrete poles of the linear density response, a considerable simplification can be achieved (Gross and Kohn 1985; Grabo et al. 2000; Casida 1995; Petersilka et al. 1996; Jamorski et al. 1996), leading to a generalized eigenvalue equation:

$$\begin{bmatrix} \mathbb{A} & \mathbb{B} \\ \mathbb{B}^* & \mathbb{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix} = \omega_n \begin{bmatrix} \mathbb{I} & 0 \\ 0 & -\mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{bmatrix}. \tag{43}$$

where the matrices $\mathbb{A}(\omega)$ and $\mathbb{B}(\omega)$ are given by

$$A_{ia\sigma,jb\tau}(\omega) = \delta_{\sigma,\tau}\delta_{i,j}\delta_{a,b}(\varepsilon_{a\sigma} - \varepsilon_{i\sigma}) + K_{ia\sigma,jb\tau}(\omega) \tag{44}$$

$$B_{ia\sigma,jb\tau}(\omega) = K_{ia\sigma,bj\tau}(\omega). \tag{45}$$

The matrices $\mathbb{A}$ and $\mathbb{B}$ are frequency-independent within the *adiabatic approximation*, which approximates that the exchange-correlation kernel $f_{xc}$ has a frequency-

independent term (Casida 2009). (Note that memory-dependent functionals were proposed (Dobson et al. 1997; Ullrich and Tokatly 2006; Wijewardane and Ullrich 2008; Kurzweil and Baer 2004), even if not commonly used.) Solving the Casida equation Eq. (43) provides excitation energies and oscillator strengths for a molecular system.

A common approximation, the Tamm-Dancoff approximation (TDA), consists in neglecting the hole-particle terms, $\mathbf{Y}_n \equiv 0$, leading to the simpler eigenvalue equation (Hirata and Head-Gordon 1999):

$$\mathbb{A}\mathbf{X}_n = \omega_n \mathbf{X}_n. \tag{46}$$

While the TDA allows for the design of better-converging algorithms (Hirata and Head-Gordon 1999; Hutter 2003), it also sometimes leads to better results than the full Casida equation (Casida et al. 2000; Tapavicza et al. 2008; Casida and Huix-Rotllant 2012). This observation might find its source from the form of the Casida equation for pure density functional theory. The Casida equation involves the linear response of the one-body density matrix and therefore accommodates the response treatment of hybrid functions in a natural way. However, when functionals with no Hartree-Fock contribution are considered, the matrix $(\mathbb{A} - \mathbb{B})$ becomes diagonal (with no Hartree-Fock exchange contribution in the functional, $(\mathbb{A} - \mathbb{B})$ becomes (Casida 2009): $(\mathbb{A} - \mathbb{B})_{ia\sigma, jb\tau} = \delta_{i,j}\delta_{a,b}\delta_{\sigma,\tau}(\varepsilon_{a\tau} - \varepsilon_{i\tau})$). Then, the exact secular equation takes a similar form as within the TDA, with $\mathbb{A}$ corrected by a contribution from $\mathbb{B}$ (Casida 2009) and relates to the exact equation derived from pure density functional response theory (Grabo et al. 2000). It is, however, important to note that within TDA the Thomas-Reiche-Kuhn sum rule is not fulfilled (Furche 2001; Hutter 2003).

### 3.2.1 Pitfalls of the Approximation of Practical LR-TDDFT

LR-TDDFT has been successfully applied to compute excitation energies and properties for a large number of molecular systems (Stratmann et al. 1998; Hirata and Head-Gordon 1999; Maitra et al. 2003; Dreuw and Head-Gordon 2005; Ullrich 2012; Casida 2009; Elliott et al. 2009; Casida and Huix-Rotllant 2012; Adamo and Jacquemin 2013; Laurent and Jacquemin 2013) However, while the LR-TDDFT formalism is *in principle* exact, its practical application to compute excitation energies for molecules requires the use of a series of approximations of the *xc*-functional and its functional derivatives (like the adiabatic approximation), which can lead to dramatic failures (Ullrich 2012; Marques et al. 2012; Casida 2009; Casida and Huix-Rotllant 2012). As a result of the adiabatic approximation, LR-TDDFT is, for example, not able to properly describe electronic states with a dominant (>50%, see Tozer and Handy 2000; Ullrich 2012) double excitation character (Hsu et al. 2001; Maitra et al. 2004; Cave et al. 2004; Levine et al. 2006; Elliott et al. 2011). Also, the combination of an inaccurate description of derivative discontinuities, the problem of self-interaction error, the incorrect long-range properties of currently used *xc*-potentials, and the adiabatic approximation are all at the heart of the most critical issue of LR-TDDFT: the charge transfer

failure (Dreuw et al. 2003; Tozer 2003; Gritsenko and Baerends 2004; Dreuw and Head-Gordon 2004; Maitra 2005; Wiggins et al. 2009; Hellgren and Gross 2012). LR-TDDFT, within the adiabatic approximation and using standard functionals, suffers to describe charge transfer excitations, i.e., excitations between a donor and an acceptor that are spatially separated. Long-range corrected functionals (Leininger et al. 1997; Iikura et al. 2001; Yanai et al. 2004) can, however, strongly improve the situation. The adiabatic approximation also leads to difficulties in describing conical intersections between the ground and first electronic state (Levine et al. 2006), even if, at least in some cases, the use of the TDA improves the description of these critical points (Tapavicza et al. 2008; Marques et al. 2012).

## 3.3    Nonadiabatic Coupling Vectors and Nuclear Forces Within LR-TDDFT

The Casida equation introduced above gives a direct access to excitation energies and oscillator strength. Nonadiabatic dynamics will require additional quantities like nonadiabatic coupling vectors (last electronic term in Eq. (8)) or excited-state nuclear forces. In the following, we will describe a strategy to compute matrix elements of one-body operator within a LR-TDDFT framework, using the concept of *auxiliary many-electron wavefunctions* that will give us access to nonadiabatic coupling vectors as well as other quantities.

### 3.3.1    Matrix Elements in LR-TDDFT

Our goal is to find a general strategy for evaluating matrix elements of the form

$$\langle \varphi_{\mathbf{R}}^{(0)} | \hat{\mathscr{O}} | \varphi_{\mathbf{R}}^{(n)} \rangle \tag{47}$$

within LR-TDDFT, where the states $|\varphi_{\mathbf{R}}^{(0)}\rangle$ and $|\varphi_{\mathbf{R}}^{(n)}\rangle$ describe the ground-state and $n$th electronic excited-state wavefunctions, respectively. To achieve this goal, we will proceed by a direct comparison with the same quantity derived using many-body perturbation theory (MBPT). Therefore, we start with a short outline of the main linear-response equations in MBPT.

From the definition of the retarded density-density response function

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \Pi^R(\mathbf{r}, t, \mathbf{r}', t') = -i\theta(t - t') \frac{\langle \varphi_{\mathbf{R}}^{(0)} | [\hat{\rho}(\mathbf{r}, t), \hat{\rho}(\mathbf{r}', t')] | \varphi_{\mathbf{R}}^{(0)} \rangle}{\langle \varphi_{\mathbf{R}}^{(0)} | \varphi_{\mathbf{R}}^{(0)} \rangle}, \tag{48}$$

the change of an observable $\mathscr{O}$, under the influence of a perturbation $v_{\text{ext}}(\mathbf{r}', t')$ in the linear-response regime, is given by

$$\delta \mathscr{O}(t) = \int_0^\infty dt' \int d\mathbf{r} \int d\mathbf{r}' \, o(\mathbf{r}) v_{\text{ext}}(\mathbf{r}', t') \chi(\mathbf{r}, t, \mathbf{r}', t') \tag{49}$$

(here we consider an interaction of the form $\delta v^{ext}(\mathbf{r}', t') = v'(\mathbf{r}')E(t')$). If $\chi$ depends only on the difference $(t - t')$, the Fourier transform in time gives

$$\delta \mathscr{O}(\omega) = \int d\mathbf{r} \int d\mathbf{r}' \, o(\mathbf{r})v'(\mathbf{r}')E(\omega)\chi(\mathbf{r}, \mathbf{r}', \omega) . \tag{50}$$

This expression can be rewritten, after a bit of algebra (Curchod et al. 2013), as a sum-over-states (SOS) formula:

$$\delta \mathscr{O}(\omega) = -2 \sum_n \frac{\omega_n \langle \varphi_{\mathbf{R}}^{(0)} | \hat{\mathscr{O}} | \varphi_{\mathbf{R}}^{(n)} \rangle \langle \varphi_{\mathbf{R}}^{(n)} | \hat{v}' E(\omega) | \varphi_{\mathbf{R}}^{(0)} \rangle}{\omega_n^2 - \omega^2} , \tag{51}$$

where $|\varphi_{\mathbf{R}}^{(n)}\rangle$ and $\omega_n$ are the true excitation energies and wavefunctions.

Meanwhile, if we use the KS representation of LR-TDDFT as above, the change of observable is in matrix representation:

$$\delta \mathscr{O}(\omega) = \sum_{ij\sigma, kl\tau} o_{ij\sigma} \chi_{ij\sigma, kl\tau}(\omega) v'_{kl\tau} E(\omega) , \tag{52}$$

where $o_{ij\sigma} = \langle \phi_{i\sigma} | \hat{\mathscr{O}} | \phi_{j\sigma} \rangle$ and $v'_{kl\tau} = \langle \phi_{l\tau} | v'(\mathbf{r}) | \phi_{k\tau} \rangle$. Similarly, a SOS formula can also be derived for LR-TDDFT (see Curchod et al. 2013 for a derivation) and reads

$$\delta \mathscr{O}(\omega) = -2 \sum_n \mathbf{o}^\dagger \frac{(\mathbb{A} - \mathbb{B})^{1/2} \mathbf{Z}_n \mathbf{Z}_n^\dagger (\mathbb{A} - \mathbb{B})^{1/2}}{\omega_n^2 - \omega^2} \mathbf{v}' E(\omega) . \tag{53}$$

with $\mathbf{Z}_n$ is related to the eigenvectors of Eq. (43) according to Casida (2009) $\mathbf{Z}_n = (\mathbb{A} - \mathbb{B})^{-1/2}(\mathbf{X}_n + \mathbf{Y}_n)$.

Comparing the residues of LR-TDDFT response function Eq. (53) with the residues of the MBPT response function Eq. (51) at equal energy $\omega_n$, we obtain the following identity:

$$\langle \varphi_{\mathbf{R}}^{(0)} | \hat{\mathscr{O}} | \varphi_{\mathbf{R}}^{(n)} \rangle = \sum_{ij\sigma}^{(f_{i\sigma} - f_{j\sigma}) > 0} \frac{1}{\sqrt{\omega_n}} o_{ij\sigma} \left( (\mathbb{A} - \mathbb{B})^{1/2} \mathbf{Z}_n \right)_{ij\sigma} . \tag{54}$$

(As stated before, with no Hartree-Fock exchange contribution in the functional, $(\mathbb{A} - \mathbb{B})$ is diagonal and becomes (Casida 2009): $(\mathbb{A} - \mathbb{B})_{ia\sigma, jb\tau} = \delta_{i,j} \delta_{a,b} \delta_{\sigma,\tau} (\varepsilon_{a\tau} - \varepsilon_{i\tau})$.)

This equation was derived by Casida (1995) and then applied by Tavernelli et al. and Hu et al. for the calculation of the nonadiabatic coupling vectors between the ground state and an excited state. A similar equation was also given in Chernyak and Mukamel (1996).

### 3.3.2 The Concept of Auxiliary Many-Electron Wavefunction

It may be useful at this point to investigate the possibility to further simplify the definition and the calculation of matrix elements within LR-TDDFT by means of the definition of a set of "auxiliary" multideterminantal many-electron wavefunctions based on KS orbitals. This route was first explored by Casida (1995) to solve the assignment problem of the LR-TDDFT excited-state transitions and then further developed by Tapavicza et al. (2007) in relation to the calculation of matrix elements in the linear and second-order response regimes (Tavernelli et al. 2009a, b, 2010).

In Tavernelli et al. (2009a), it was showed that defining the ground-state many-electron wavefunction $\langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{N_{el}} | \tilde{\varphi}_{\mathbf{R}}^{(0)} \rangle$ as a Slater determinant of all occupied KS orbitals $\{\phi_i\}_{i=1}^{N_{el}}$ and the excited-state wavefunction corresponding to the excitation energy $\omega_n$ as

$$\langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{N_{el}} | \tilde{\varphi}_{\mathbf{R}}^{(n)} \rangle = \sum_{ia\sigma} \sqrt{\frac{\varepsilon_a - \varepsilon_i}{\omega_n}} (\mathbf{Z}_n)_{ia\sigma} \hat{a}_{a\sigma}^\dagger \hat{a}_{i\sigma} \langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{N_{el}} | \tilde{\varphi}_{\mathbf{R}}^{(0)} \rangle$$

$$= \sum_{ia\sigma} \mathscr{C}_{ia\sigma}^n \langle \mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{N_{el}} | \tilde{\varphi}_{\mathbf{R},i\sigma}^{a\sigma} \rangle, \tag{55}$$

we obtain for any one-body operator of the form $\hat{\mathscr{O}} = \sum_{pq\sigma} o_{pq\sigma} \hat{a}_{p\sigma}^\dagger \hat{a}_{q\sigma}$ (where $p, q$ are general indices) the correct linear-response expression for the matrix element $\langle \varphi_{\mathbf{R}}^{(0)} | \hat{\mathscr{O}} | \varphi_{\mathbf{R}}^{(n)} \rangle$. Eq. (55) is derived from Eq. (54) where now the index $i$ runs over all occupied and $a$ over the unoccupied (virtual) KS orbitals and $|\tilde{\varphi}_{\mathbf{R},i\sigma}^{a\sigma}\rangle$ denotes a singly excited Slater determinant defined by the transition $i\sigma \rightarrow a\sigma$. This theory was then successfully extended to the case of the calculation of matrix elements between two excited-state wavefunctions, $\langle \varphi_{\mathbf{R}}^{(n)} | \hat{\mathscr{O}} | \varphi_{\mathbf{R}}^{(m)} \rangle$ as will be briefly discussed in the next section on the calculation of nonadiabatic coupling vectors.

It is important to further stress the fact that both auxiliary functions introduced above have a physical meaning only when used within LR-TDDFT for the calculation of matrix elements of the type $\langle \tilde{\varphi}_{\mathbf{R}}^{(0)} | \hat{\mathscr{O}} | \tilde{\varphi}_{\mathbf{R}}^{(n)} \rangle$ and eventually $\langle \tilde{\varphi}_{\mathbf{R}}^{(n)} | \hat{\mathscr{O}} | \tilde{\varphi}_{\mathbf{R}}^{(m)} \rangle$. The use of this representations of the ground-state and excited-state KS many-electron wavefunctions in other contexts is not justified.

### 3.3.3 Nonadiabatic Coupling Vectors Within LR-TDDFT

Using the concept of the auxiliary many-electron wavefunction approach described above, we can now propose an approach for the calculation of nonadiabatic vectors within LR-TDDFT.

#### Couplings Between Ground and Excited States

One can use an alternative definition of the nonadiabatic coupling vectors (Epstein 1954) (see also Ch. 5 of Baer 2006 for a complete discussion) between the ground (0) state and the $n^{th}$ excited state for a molecular system characterized by nuclear coordinates $\mathbf{R}$ in the configuration space ($\mathbb{R}^{3N_n}$):

$$\mathbf{d}_{0n}^{\gamma} = \frac{\langle \varphi_{\mathbf{R}}^{(0)} | \partial_{\gamma} \hat{H}_{BO} | \varphi_{\mathbf{R}}^{(n)} \rangle}{\varepsilon_{BO}^{(n)}(\mathbf{R}) - \varepsilon_{BO}^{(0)}(\mathbf{R})} \tag{56}$$

where $\gamma$ is an atomic label, $\hat{H}_{BO}$ is the electronic Hamiltonian, and $\partial_{\gamma} \hat{H}_{BO} = \partial \hat{H}_{BO} / \partial \mathbf{R}_{\gamma}$.

Applying the results of the above sections on the evaluation of matrix elements of the form $\langle \varphi_{\mathbf{R}}^{(0)} | \hat{\mathscr{O}} | \varphi_{\mathbf{R}}^{(n)} \rangle$ in LR-TDDFT to the nonadiabatic coupling vectors gives directly the expression:

$$\mathbf{d}_{0n}^{\gamma} = \sum_{ij\sigma}^{(f_{i\sigma} - f_{j\sigma})>0} \frac{1}{(\omega_n)^{3/2}} h_{ij\sigma}^{\gamma} \left( (\mathbb{A} - \mathbb{B})^{1/2} \mathbf{Z}_n \right)_{ij\sigma} \tag{57}$$

where $h_{ij\sigma}^{\gamma} = \int d\mathbf{r} \, \partial_{\gamma} \hat{H}_{BO} \, \phi_{i\sigma}^{*}(\mathbf{r}) \phi_{j\sigma}(\mathbf{r})$.

This formula for the nonadiabatic coupling vectors within LR-TDDFT was derived several times in the literature using slightly different formalisms. The first derivation was given by Chernyak and Mukamel (2000) using a classical Liouville dynamics for the single-electron density matrix, followed by Baer (2002). Later, Tapavicza et al. (2007), Tavernelli et al. (2009b), and Hu et al. (2007, 2008) arrived to the same result (Eq. (57)) using the formulation based on Casida's LR-TDDFT equations (Casida 1995).

Concerning the numerical implementation of Eq. (56), several approaches have also been proposed that differ mainly in the choice of the basis set and in the way the implicit dependence of the pseudopotentials on the nuclear positions is treated. Due to the technical nature of this subject, we will not go through the numerical details but better refer to the literature, which is very rich on this subject (Tavernelli et al. 2009b; Hu et al. 2007, 2010, 2012; Send and Furche 2010).

### Couplings Between Excited States

LR-TDDFT only gives access strictly speaking to the couplings between ground and excited state. However, the concept of LR-TDDFT auxiliary many-electron wavefunctions can also be used as a good approximation, exact within the Tamm-Dancoff approximation, to compute couplings between excited states (Tavernelli et al. 2010), $\mathbf{d}_{kn}$. An exact derivation of these coupling terms *beyond* the linear-response formalism of TDDFT was also proposed in the literature (Li and Liu 2014; Li et al. 2014; Ou et al. 2015). However, this formalism implies the calculation of an exchange-correlation hyperkernel and leads to the critical appearance of divergences in the couplings as a result of the adiabatic approximations (Parker et al. 2016).

## 3.4    Nuclear Forces Within LR-TDDFT

Excited-state dynamics using LR-TDDFT will also require the calculation of nuclear forces. Among the different approaches developed for the calculation of analytic derivatives, the Lagrangian method (Helgaker and Jørgensen 1989) is of particular interest because of its numerical efficiency. However, the derivation of LR-TDDFT forces is technically involved and goes beyond the scope of this chapter. We refer the interested reader to the abundant literature on the subject (Pulay 1987; Hutter 2003; Deglmann et al. 2002; Rappoport and Furche 2005; Marx and Hutter 2009).

## 4    Nuclear Dynamics: Trajectory-Based Quantum-Classical Dynamics

In this section, different approaches to nonadiabatic electron-nuclear dynamics will be presented, namely, the Ehrenfest scheme (Tully 1998), surface hopping (Tully 1990), the coupled-trajectory mixed quantum-classical (CT-MQC) method derived from the exact factorization (Min et al. 2015), and full multiple spawning (Martínez et al. 1996). Their common feature is the use of trajectories to explore the nuclear configuration space, which are subject to the time-dependent effect of the electrons in the ground state as well as in the excited states. The electronic properties needed in the calculations can be determined on-the-fly based on ab initio electronic structure methods. For the purpose of this work, TDDFT and its LR formulation will be employed. Other approaches based on Bohmian trajectories are also possible (Curchod et al. 2011; Curchod and Tavernelli 2013; Tavernelli 2013), but they will not be discussed in this book chapter.

Ehrenfest, surface hopping, and CT-MQC are based on a purely classical description of nuclear motion, which is coupled to the quantum-mechanical evolution of the electrons. In the three approaches, a hypothesis is made to decompose the full TDSE into two coupled equations, one describing the evolution of the electronic subsystem and the other describing the evolution of the nuclear subsystem. The main difference among them lies in the procedure followed for such decomposition. In particular, only the exact factorization starts from an Ansatz for the molecular wavefunction, which translates into exact coupled electronic and nuclear equations. Only in a second step, the nuclear evolution is modeled using classical trajectories. The full multiple spawning scheme, on the other hand, introduces an expansion in terms of Gaussian wave packets to represent each nuclear coefficients $\chi_k(\mathbf{R}, t)$ of the Born-Huang expansion (4). The parameters of the Gaussians are evolved classically, under the assumption that classical dynamics samples correctly the nuclear configuration space. Indeed, in the limit of an infinite number of Gaussians, full multiple spawning converges to an exact description of the electron-nuclear problem.

## 4.1 Ehrenfest Dynamics

To derive Ehrenfest decomposition, one makes the assumption that the full wave-function can be written as a single product of a purely electronic $\Phi(\mathbf{r}, t)$ and a purely nuclear $\chi(\mathbf{R}, t)$ wavefunction:

$$\Psi(\mathbf{r}, \mathbf{R}, t) = e^{\frac{i}{\hbar} \int_0^t dt' E_{BO}(t')} \Phi(\mathbf{r}, t) \chi(\mathbf{R}, t) . \tag{58}$$

Here, the time-dependent phase on the right-hand side is inserted to simplify the following equations derived from such an Ansatz; thus the energy $E_{BO}(t)$ is chosen as

$$E_{BO}(t) = \int d\mathbf{r} \, \Phi^*(\mathbf{r}, t) i\hbar \partial_t \Phi(\mathbf{r}, t) . \tag{59}$$

The product form of the molecular wavefunction in Eq. (58) is clearly uncorrelated, and in this initial Ansatz lies the fundamental approximation of the Ehrenfest scheme. When Eq. (58) is inserted into the molecular TDSE (1), the coupled equations

$$\left[ \hat{T}_e(\mathbf{r}) + \hat{V}_{ee}(\mathbf{r}) + \int d\mathbf{R} \, \chi^*(\mathbf{R}, t) \left[ \hat{V}_{nn}(\mathbf{R}) + \hat{V}_{en}(\mathbf{r}, \mathbf{R}) \right] \chi(\mathbf{R}, t) \right] \Phi(\mathbf{r}, t)$$
$$= i\hbar \partial_t \Phi(\mathbf{r}, t) \tag{60}$$

$$\left[ \sum_v \frac{-\hbar^2}{2M_v} \nabla_v^2 + \int d\mathbf{r} \, \Phi^*(\mathbf{r}, t) \hat{H}_{BO}(\mathbf{r}, \mathbf{R}) \Phi(\mathbf{r}, t) \right] \chi(\mathbf{R}, t)$$
$$= i\hbar \partial_t \chi(\mathbf{R}, t) \tag{61}$$

are derived, by averaging over the instantaneous nuclear, in Eq. (60), and electronic, in Eq. (61), state. In both equations, the wavefunctions $\Phi(\mathbf{r}, t)$ and $\chi(\mathbf{R}, t)$ are supposed to be normalized. Therefore, Eq. (60) describes the evolution of the electrons in the mean field created by the nuclei, whereas the nuclei move according to Eq. (61) in the mean field of the electrons.

A quantum-classical algorithm can be derived from Eqs. (60) and (61) by approximating classically the nuclear equation, that is, by determining the force to propagate the nuclei as trajectories. A standard procedure can be followed, by introducing a complex-phase representation of $\chi(\mathbf{R}, t)$ and by only considering terms $\mathcal{O}(\hbar^0)$ in the asymptotic expansion of the complex phase in powers of $\hbar$ (Van Vleck 1928). The equation for the zeroth order term $S(\mathbf{R}, t)$ of this expansion is thus obtained, namely:

$$\partial_t S(\mathbf{R}, t) = - \left[ \sum_v \frac{[\nabla_v S(\mathbf{R}, t)]^2}{2M_v} + \int d\mathbf{r} \, \Phi^*(\mathbf{r}, t) \hat{H}_{BO}(\mathbf{r}, \mathbf{R}) \Phi(\mathbf{r}, t) \right] . \tag{62}$$

This Hamilton-Jacobi-like equation can be solved via characteristics, thus yielding the expression of the classical (Ehrenfest) force as

$$\mathbf{F}_v(t) = -\nabla_v \int d\mathbf{r}\, \Phi^*(\mathbf{r}, t)\hat{H}_{BO}(\mathbf{r}, \mathbf{R})\Phi(\mathbf{r}, t)\,. \tag{63}$$

The classical approximation is also introduced in the electronic evolution equation (60). Here, the nuclear density $|\chi(\mathbf{R}, t)|^2$ is approximated as a product of $\delta$-functions centered at each time at the position of the classical nuclei, that is, at the position of the classical trajectory $\mathbf{R}^{(I)}(t)$. Therefore, the TDSE describing the evolution of $\Phi(\mathbf{r}, t)$ becomes

$$\hat{H}_{BO}\left(\mathbf{r}, \mathbf{R}^{(I)}(t)\right)\Phi\left(\mathbf{r}, \mathbf{R}^{(I)}(t), t\right) = i\hbar\partial_t\Phi\left(\mathbf{r}, \mathbf{R}^{(I)}(t), t\right). \tag{64}$$

The electronic wavefunction acquires an implicit dependence on the nuclear positions, expressed as the classical trajectory, via the dependence of the BO Hamiltonian on $\mathbf{R}^{(I)}(t)$. The trajectory $I$ of the nucleus $v$ is determined by solving Newton's equation with force:

$$\mathbf{F}_v^{(I)}(t) = \int d\mathbf{r}\, \Phi^*(\mathbf{r}, \mathbf{R}^{(I)}(t), t)\left[-\nabla_v\hat{H}_{BO}(\mathbf{r}, \mathbf{R}^{(I)}(t))\right]\Phi(\mathbf{r}, \mathbf{R}^{(I)}(t), t)\,, \tag{65}$$

where now a label $(I)$ has been introduced to show that along a trajectory, Eqs. (64) and (65) have to be evolved consistently. Multiple trajectories can also be employed, to "wash out" some of the details of the coherent evolution along a single trajectory. Nuclear and electronic observables can thus be determined as averages over this ensemble of trajectories.

As it provides the true time-dependent electronic density, TDDFT can be used within an Ehrenfest dynamics scheme to perform nonadiabatic molecular dynamics. The mapping of the nuclear equation (Eq. (65)) into the DFT formalism is straightforward and only requires the description of the forces $\langle -\nabla_v\hat{H}_{BO}(\mathbf{r}, \mathbf{R}^{(I)}(t))\rangle$ as a functional of the time-dependent density $\rho(\mathbf{r}, t)$. If we replace the expectation value of the electronic Hamiltonian with the DFT energy evaluated with the exchange-correlation potential $v_{xc}[\rho]|_{\rho(\mathbf{r})\leftarrow\rho(\mathbf{r},t)}$, the gradient with respect to the nuclear coordinates can be performed analytically as in the case of the adiabatic BO dynamics and the Car-Parrinello (Car and Parrinello 1985) molecular dynamics schemes (Marx and Hutter 2009).

### 4.1.1 Application of Ehrenfest Dynamics Combined with TDDFT

As Ehrenfest dynamics gives a direct access to electronic dynamics, it is a method of choice to investigate the dynamics of the electronic density and subsequent nuclear dynamics after a strong perturbation. Such perturbation can be induced by the action of an external light pulse or through the collision with a highly charged particle, generating either an electronic excitation or, in the some other cases, electron abstraction (Tavernelli et al. 2005; Tavernelli 2006, 2015; Castro

et al. 2004; Li et al. 2005; Yagi and Takatsuka 2005; Andrade et al. 2009; Moss et al. 2009; Liang et al. 2010; Gaigeot et al. 2010; Lopez-Tarifa et al. 2011; Elliott and Maitra 2012). The latter takes place when an XUV attosecond pulse interacts with a molecule and leads to a core ionization. In a Born-Huang picture, such an ultrafast ionization leads to the generation of an electronic wave packet, i.e., the generation of a coherent superposition of different nuclear contributions on a large number of electronic states (the number of electronic states being considered depends on the bandwidth of the ionizing pulse). Ehrenfest dynamics combined with TDDFT offers an alternative to the Born-Huang picture by only requiring the generation of an initial electronic density to represent the initial ionized state. Martín et al. employed this strategy to study the role of nuclear motion in the electronic dynamics upon XUV ultrafast ionization of a small amino acid, glycine (Lara-Astiaso et al. 2017). The one-electron ionization generated by the sub-300-as XUV pulse generates an electronic wave packet that can be described by a coherent superposition of more than ten one-hole states, in an energy domain ranging from 17 to 35 eV (the pulse bandwidth). The electronic density corresponding to this electronic wave packet, $\rho(\mathbf{r}, t_0)$, is used as initial condition for two simulations: (*i*) real-time TDDFT with frozen nuclei and (*ii*) real-time TDDFT combined with Ehrenfest dynamics. As a result of the nature of the electronic wave packet, the time evolution of the unpaired electron (with respect to the initial density) shows that the electron migrates over the entire molecular scaffold with a dynamics that is characterized by only few, system dependent, frequencies (Fig. 1). This is the behavior that one would expect in the case the ionized electron is removed from one given localized orbital. Comparing the two panels of Fig. 1, one can observe that nuclear motion starts altering the electronic dynamics already after the first 10 fs of dynamics, emphasizing the importance of including nuclear dynamics in such simulations. However, it is important to note that the mean-field character of Ehrenfest dynamics might hamper a more detailed study of the electronic wave packet dynamics, in particular due to the underestimation of decoherence and dephasing effects at longer timescales (Vacher et al. 2017). The ease of the Ehrenfest formalism combined with
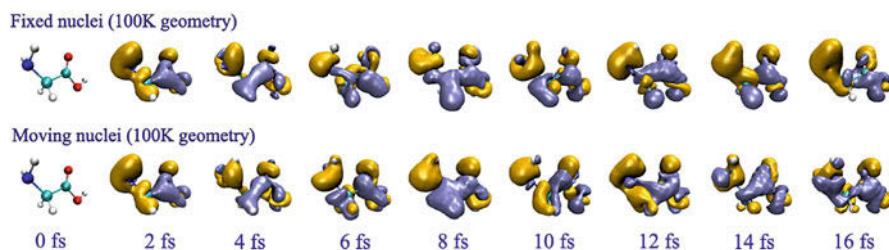


**Fig. 1** Spin density differences at different times after interaction of a XUV attosecond pulse with the glycine molecule. The initial conditions correspond to a geometry obtained after thermalization at 100 K. (Adapted from Lara-Astiaso et al. 2017, Copyright (2017), with permission from Elsevier)

the efficiency of TDDFT offers nevertheless a valid tool for the study of the short-time electronic wave packet dynamics in molecular systems.

## 4.2 Surface Hopping

Surface-hopping decomposition is derived under the preliminary assumption that the nuclei evolve along classical trajectories. Therefore, the BO Hamiltonian acquires an implicit time dependence via its dependence on the nuclear coordinates. A TDSE is proposed in this way, namely:

$$\hat{H}_{BO}(\mathbf{R}^{(I)}(t))\Phi_{\mathbf{R}^{(I)}(t)}(\mathbf{r}, t) = i\hbar\partial_t\Phi_{\mathbf{R}^{(I)}(t)}(\mathbf{r}, t), \tag{66}$$

for the electronic wavefunction, which is, itself, dependent on the classical trajectories. As done above, a classical trajectory is labeled by the index $(I)$. An expansion in the adiabatic basis is introduced for $\Phi_{\mathbf{R}^{(I)}(t)}(\mathbf{r}, t)$:

$$\Phi_{\mathbf{R}^{(I)}(t)}(\mathbf{r}, t) = \sum_k C_k(t)\varphi^{(k)}_{\mathbf{R}^{(I)}(t)}(\mathbf{r}), \tag{67}$$

and Eq. (66) yields

$$\dot{C}^{(I)}_k(t) = \frac{-i}{\hbar}\varepsilon^{(k)}_{BO}\left(\mathbf{R}^{(I)}(t)\right)C^{(I)}_l(t) - \sum_l C^{(I)}_l(t)\sum_{v=1}^{N_n}\frac{\mathbf{P}^{(I)}_v(t)}{M_v}\cdot\mathbf{d}_{v,kl}\left(\mathbf{R}^{(I)}(t)\right). \tag{68}$$

Here, the BO PES $\varepsilon^{(k)}_{BO}(\mathbf{R})$ and the nonadiabatic coupling vectors, i.e.,

$$\mathbf{d}_{v,kl}(\mathbf{R}) = \int d\mathbf{r}\,\varphi^{(k)*}_{\mathbf{R}}(\mathbf{r})\nabla_v\varphi^{(l)}_{\mathbf{R}}(\mathbf{r}) = \left\langle\varphi^{(k)}_{\mathbf{R}}\middle|\nabla_v\varphi^{(l)}_{\mathbf{R}}\right\rangle_{\mathbf{r}}, \tag{69}$$

which are functions of the nuclear coordinates, are evaluated at the instantaneous positions along the trajectories; they thus become functions of the trajectory itself. Henceforth, a superscript $(I)$ will be introduced to indicate this dependence on the trajectory.

The surface-hopping scheme takes its name from the idea suggested for the evolution of the classical nuclear trajectories, namely, that a trajectory evolves according to one adiabatic BO force, determined as (minus) the gradient of the BO potential energy surface (PES), until a stochastic hop occurs onto another BO PES. The classical (surface-hopping) force can then be written as

$$\mathbf{F}^{(I)}_v(t) = -\nabla_v\varepsilon^*_{BO}, \tag{70}$$

with the symbol * indicating that the force state is selected stochastically at each time step. The discontinuity in the force, and thus in the potential energy, for a given trajectory, is compensated by a discontinuity in the velocity, and thus in the kinetic energy, that guarantees energy conservation. The hopping scheme fewest switches (Tully 1990) prescribes that the trajectory $I$ hops from surface $k$ to surface $l$ according to the probability:

$$\mathscr{P}_{k \rightarrow l} = \max \left[ 0, \frac{-2\,dt}{\left| C_k^{(I)}(t) \right|^2} \; \text{Re} \left[ C_k^{(I)^*}(t) C_l^{(I)}(t) \right] \sum_v \frac{\mathbf{P}_v^{(I)}(t)}{M_v} \cdot \mathbf{d}_{lk,v}^{(I)} \right], \quad (71)$$

with $dt$ the integration time step.

The major drawback of the surface-hopping scheme is the (over)coherent evolution of the electronic coefficients coupled to the classical (independent) trajectories. The issue has been well-documented in the literature (Subotnik et al. 2013; Bittner and Rossky 1995; Curchod and Tavernelli 2013; Gao and Thiel 2017), and several schemes have been proposed (Shenvi et al. 2011a, b; Shenvi and Yang 2012; Subotnik and Shenvi 2011a, b; Jaeger et al. 2012; Jasper and Truhlar 2007; Granucci and Persico 2007) to cure or alleviate this shortcoming.

### 4.2.1   Application of Surface Hopping Combined with LR-TDDFT

Surface hopping has been used to study a large number of excited-state mechanisms, and we refer the interested reader to specialized reviews (Barbatti 2011; Curchod et al. 2013; Persico and Granucci 2014) for a list of applications. The application presented here highlights the combination of surface hopping with LR-TDDFT (using the concepts developed in Sect. 3.2), including implicitly and explicitly the role of spin-orbit coupling as well as explicit solvent effects. Ruthenium (II) trisbipyridine, $[\text{Ru(bpy)}_3]^{2+}$, is an inorganic molecule recognized for its extremely efficient intersystem crossing process, i.e., when the molecule changes, in this particular case, from a singlet electronic state to a triplet electronic state (Cannizzo et al. 2006; Gawelda et al. 2006). $[\text{Ru(bpy)}_3]^{2+}$ is initially photoexcited in a singlet metal-to-ligand-charge-transfer ($^1$MLCT) state before it rapidly relaxes among other $^1$MLCT or $^3$MLCT, as a result of the high density of states; the overall dynamics to the triplet states has been observed experimentally in water within a ∼50 fs timescale (Fig. 2).

In the first theoretical study (Tavernelli et al. 2011), the excited-state dynamics of the $[\text{Ru(bpy)}_3]^{2+}$ in water was studied by employing surface hopping with LR-TDDFT, in a QM/MM formalism where water molecules were treated classically. Intersystem-crossing events were analyzed a posteriori, monitoring the crossings between singlet and triplet states and evaluating spin-orbit coupling from qualitative rules. Owing to the cost of the overall dynamics, this study was limited to only two trajectories. Nevertheless, both trajectories indicated an ultrafast decay of the molecule toward triplet states in less than 50 fs, in good correlation with experimental evidences. The MLCT character of the different excited states implies
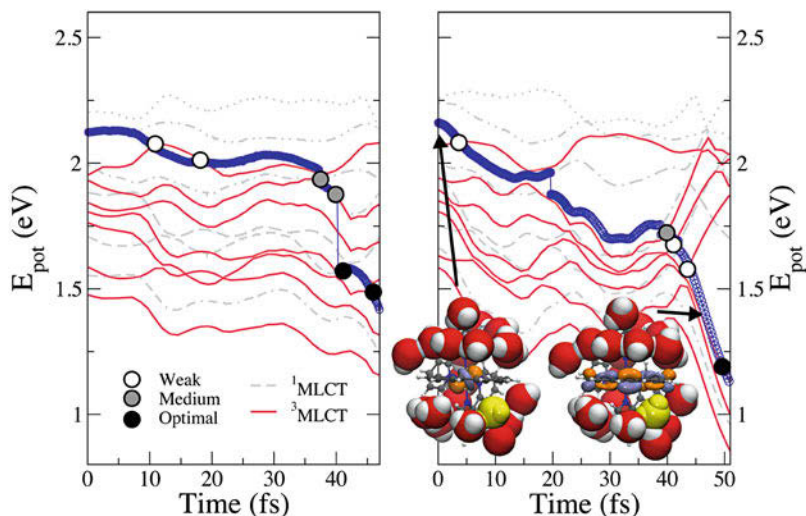
**Fig. 2** Two surface-hopping trajectories for $[Ru(bpy)_3]^{2+}$ in water. The driving state is highlighted with blue circles, while the seven singlet excited states considered in the surface hopping dynamics are represented by gray dashed lines and the seven triplet states by red continuous lines. Filled circles indicate the analyzed crossings between singlet and triplet states, using the following color coding: white = weak, gray = medium, and black = optimal SOC strength. The inset provides a ball-and-stick representation of the $[Ru(bpy)_3]^{2+}$ molecule with part of its first water solvation shell of water molecules for two selected frames (black arrows), highlighting the fast rotation of a classical water molecule (in yellow) occurring during the dynamics. (Adapted from Tavernelli et al. 2011, Copyright (2011), with permission from Elsevier)

that an electron moves from the central metal to one (or two, depending on the state) solvent-exposed bipyridine ligands. Hence, the simulation showed that water molecules in the first solvation shell can rapidly rearrange in a non-diffusive rotation around the hydrogen bond axis to stabilize an extra charge located on a close ligand. An explicit treatment of solvent molecules is central to capture such effects as well as a proper ordering of the different electronic states.

In a more recent study (Atkins and González 2017), surface hopping combined within LR-TDDFT was used to simulate the excited-state dynamics of $[Ru(bpy)_3]^{2+}$ in gas phase but with the explicit treatment of spin-orbit coupling in a perturbative ZORA formalism (Wang and Ziegler 2005) and a larger number (101) of trajectories. This study confirmed the ultrafast decay of the original $^1$MLCT population toward triplet states, already at the early time of the dynamics. Horizontal intersystem crossing processes were observed, followed by ultrafast nonadiabatic dynamics among the triplet states (Fig. 3). Thanks to normal mode and principal component analysis, the authors could identify that the motion of both the ruthenium and the coordinated nitrogens is activated, even within such short timescale, leading potentially to the intersystem crossing events.
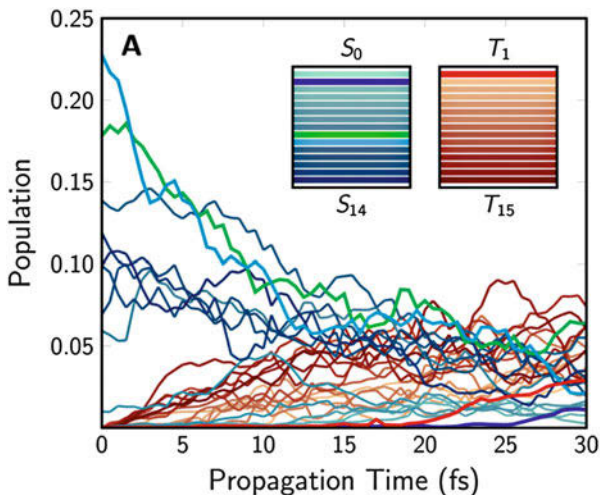
**Fig. 3** Population trace of singlet (bluish) and triplet (brownish) states over 30 fs of surface hopping (population of the three sublevels of each triplet are summed together). (Adapted with permission from Atkins and González 2017, Copyright (2017) American Chemical Society)

### 4.3    Coupled-Trajectory Mixed Quantum-Classical Scheme

Within the exact factorization formalism, a trajectory-based solution of the electronic (11) and nuclear (12) equations is constructed by (i) determining the classical limit of the nuclear equation, thus deriving the corresponding Newton's equation with forces computed from the time-dependent vector $\mathbf{A}_v(\mathbf{R}, t)$ and scalar $\varepsilon(\mathbf{R}, t)$ potentials, (ii) introducing the Born-Huang-like expansion of Eq. (16) of the electronic wavefunction, (iii) approximating the explicit dependence on the nuclear wavefunction, i.e., the term $-i\hbar\nabla_v\chi(\mathbf{R}, t)/\chi(\mathbf{R}, t)$ in the definition of the coupling operator $\hat{U}_{en}^{\text{coup}}[\Phi_{\mathbf{R}}, \chi]$ (13) employing information obtained from the trajectories. A thorough account of the steps adopted for the derivation of the algorithm is given in Agostini et al. (2016). Following this procedure, the electronic and nuclear equations of the exact factorization can be rewritten as

$$\dot{C}_l^{(I)}(t) = \dot{C}_{\text{Eh}.l}^{(I)}(t) + \dot{C}_{\text{qm}\,l}^{(I)}(t) \tag{72}$$

$$\mathbf{F}_v^{(I)}(t) = \mathbf{F}_{\text{Eh}.\,v}^{(I)}(t) + \mathbf{F}_{\text{qm}\,v}^{(I)}(t) \,. \tag{73}$$

The electronic equation yields a set of ordinary differential equations $\dot{C}_l^{(I)}(t)$ for the expansion coefficients in the Born-Huang expansion, each labeled by a superscript $(I)$ indicating that they are calculated along the $I$−th classical trajectory. The nuclear equation allows one to identify the classical force $\mathbf{F}_v^{(I)}(t)$ acting on the $v$−th nucleus that evolves along the $I$−th trajectory. Both equations can be decomposed

as the sum of two terms: the first, indicated by Eh., comprises Ehrenfest-like terms, while the second, qm, originates from the exact factorization. These last terms depend on the so-called quantum momentum, whose expression is given below. The Ehrenfest-like terms are

$$\dot{C}^{(I)}_{\text{Eh.},l}(t) = \frac{-i}{\hbar}\varepsilon^{(l)(I)}_{BO}C^{(I)}_l(t) - \sum_k C^{(I)}_k(t)\sum_{v=1}^{N_n}\frac{\mathbf{P}^{(I)}_v(t)}{M_v}\cdot\mathbf{d}^{(I)}_{v,lk} \tag{74}$$

$$\mathbf{F}^{(I)}_{\text{Eh.},v}(t) = -\sum_k\left|C^{(I)}_l(t)\right|^2\nabla_v\varepsilon^{(k),(I)}_{BO} - \sum_{k,l}C^{(I)*}_l(t)C^{(I)}_k(t)\left(\varepsilon^{(k),(I)}_{BO} - \varepsilon^{(l),(I)}_{BO}\right)\mathbf{d}^{(I)}_{v,lk}, \tag{75}$$

where we introduced the symbols $\varepsilon^{(l)(I)}_{BO}$ for the electronic adiabatic potential energy surface corresponding to state $l$ and evaluated at the position of the $I-$th trajectory, $\mathbf{d}^{(I)}_{v,lk}$ for the nonadiabatic coupling vectors defined as $\langle\varphi^{(l)(I)}|\nabla_v\varphi^{(k)(I)}\rangle_{\mathbf{r}}$, as well evaluated at the position of the trajectory $I$, and $\mathbf{P}^{(I)}_v(t)$ for the classical momentum of the $v-$th nucleus evolving along the $I-$th trajectory. The additional terms in Eqs. (72) and (73), namely:

$$\dot{C}^{(I)}_{\text{qm}\,l}(t) = -\sum_{v=1}^{N_n}\frac{\mathcal{Q}^{(I)}_v(t)}{\hbar M_v}\cdot\left[\sum_k\left|C^{(I)}_k(t)\right|^2\mathbf{f}^{(I)}_{k,v}(t) - \mathbf{f}^{(I)}_{l,v}(t)\right]C^{(I)}_l(t), \tag{76}$$

$$\mathbf{F}^{(I)}_{\text{qm}\,v}(t) = -\sum_l\left|C^{(I)}_l(t)\right|^2\left(\sum_{v'=1}^{N_n}\frac{2}{\hbar M_{v'}}\mathcal{Q}^{(I)}_{v'}(t)\cdot\mathbf{f}^{(I)}_{l,v'}(t)\right)$$
$$\left[\sum_k\left|C^{(I)}_k(t)\right|^2\mathbf{f}^{(I)}_{k,v}(t) - \mathbf{f}^{(I)}_{l,v}(t)\right], \tag{77}$$

can be derived *only* in the context of the exact factorization, as they both depend on the quantum momentum (Garashchuk and Rassolov 2003) $\mathcal{Q}^{(I)}_v(t) = -\hbar(\nabla_v|\chi^{(I)}(t)|^2)/(2|\chi^{(I)}(t)|^2)$. Here, $|\chi^{(I)}(t)|^2$ stands for the value of the nuclear density evaluated at the position of the $I-$th trajectory. The quantum momentum appears in the expression of $\hat{U}^{coup}_{en}[\Phi_{\mathbf{R}},\chi]$ as a purely imaginary correction to the (real-valued) classical momentum. As exhaustively described in Agostini et al. (2016), the evaluation of the quantum momentum along the $I-$th trajectory at a given time requires knowledge of the positions of all other trajectories at the same time. This peculiar feature couples the trajectories in a nontrivial manner, thus allowing for the correct description of quantum decoherence effects. The additional new quantities appearing in Eqs. (76) and (77) are the adiabatic forces accumulated over time $\mathbf{f}^{(I)}_{l,v}(t) = -\int^t dt'\nabla_v\varepsilon^{(l),(I)}_{BO}$.

The electronic structure inputs required in the propagation of the CT-MQC equations of motion are the Born-Oppenheimer energies and the nonadiabatic

coupling vectors. Any electronic structure methods providing these quantities can therefore be used in combination with CT-MQC. In the following, we make the choice of employing LR-TDDFT.

### 4.3.1 Applications

As an application of the CT-MQC approach, we report the analysis presented in Min et al. (2017) and Curchod et al. (2018) where the photoinduced ring-opening process in oxirane (Cordova et al. 2007; Tapavicza et al. 2008) triggered by the excitation from $S_0$ to $S_2$. The following discussion is based on Curchod et al. (2018). Electronic structure calculations are performed with the CPMD code using the GGA functional PBE (Perdew et al. 1996) for ground state and excited states. Linear-response TDDFT calculations are based on the Tamm-Dancoff approximation (Tamm 1945; Dancoff 1950). The Kleinman-Bylander (Kleinman and Bylander 1982) pseudo-potential has been used for all atom species together with a plane-wave cutoff of 70 Ry. Initial conditions, i.e., positions and momenta, have been sampled from an ab initio ground-state 2 ps trajectory at 300 K. $N_{tr} = 100$ trajectories are propagated with a time step of 0.12 fs (5 a.u.).

Figure 4 shows the electronic populations (upper panel) and an indicator of decoherence (lower panel), both quantities averaged over the 100 coupled trajectories. As indicated in Eq. (72), the coefficients in the expansion of the electronic wavefunction are labeled by the trajectory index ($I$); therefore the average population can be determined as

$$\rho_k(t) = \frac{1}{N_{tr}} \sum_{I=1}^{N_{tr}} \left| C_k^{(I)}(t) \right|^2 \quad \text{for } k = 1, 2 \,. \tag{78}$$

Between 7 and 15 fs, the trajectories cross the coupling region $S_1/S_2$; thus population is transferred from the initially occupied electronic state $S_2$ to $S_1$. After about 20 fs, the nonadiabatic event is almost complete, and some trajectories evolving in $S_1$ will encounter a second coupling region $S_0/S_1$, as it is clearly shown by the increase of population of $S_0$ (and consequent decrease of population of $S_1$) at about 25 fs.

Similarly to Eq. (78), the indicator of decoherence introduced in Agostini et al. (2016), and Min et al. (2015, 2017) is computed between the states $S_1$ and $S_2$ as an average over the trajectories:

$$\eta_{12}(t) = \frac{1}{N_{tr}} \sum_{I=1}^{N_{tr}} \left| C_1^{(I)^*}(t) C_2^{(I)}(t) \right|^2 \,. \tag{79}$$

The quantity $C_1^{(I)^*}(t) C_2^{(I)}(t)$ stands for the off-diagonal element of the electronic density matrix in the adiabatic representation between the first two excited states and depends on nuclear positions through the dependence on the trajectory index ($I$). The decoherence indicator depends on the choice of the representation used to
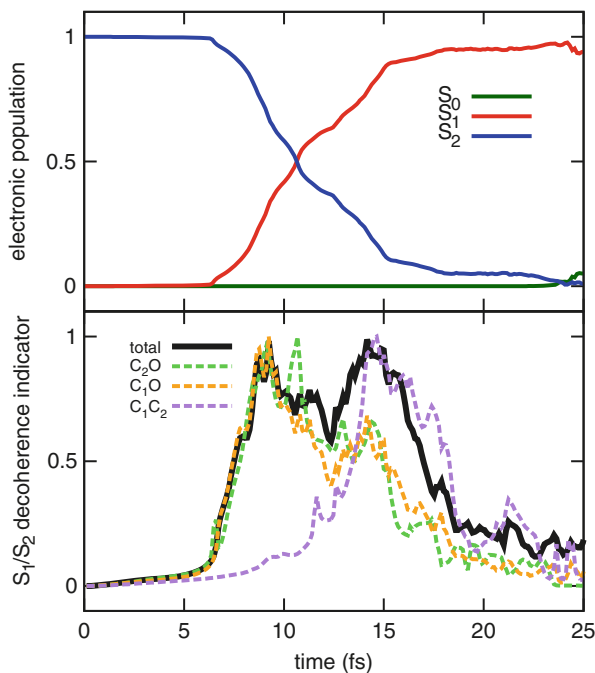
**Fig. 4** Upper panel: electronic populations of $S_0$ (dark-green line), $S_1$ (red line), and $S_2$ (blue line) as functions of time. Lower panel: (normalized) indicator of decoherence for the element $S_1/S_2$ (black line) and its decomposition contributions arising from three sets of trajectories. The sets of trajectories labeled by $C_1O$ (dashed orange line) and $C_2O$ (dashed green line) lead to a final configuration where the oxirane ring opens via the breaking of one of the two equivalent CO bonds; the set of trajectories labeled $C_1C_2$ (dashed purple line) yields final configurations where the ring opens via elongation of the CC bond

describe the electronic states. Our particular choice has fallen on the adiabatic representation, which is a natural choice since the dynamics is simulated in the adiabatic basis. Furthermore, this indicator of decoherence contains information simultaneously about electronic coherences and nuclear dynamics, via the parametric dependence of the adiabatic basis on the nuclear coordinates. Decoherence can thus be related to the spatial separation in configuration space of different bundles of trajectories (and thus of different wave packets), which "loose memory" of each other while evolving along diverging paths after funneling through the conical intersection. As abundantly discussed in the literature (Jasper et al. 2006; Granucci and Persico 2007; Jaeger et al. 2012; Subotnik et al. 2013; Gao and Thiel 2017; Schwartz et al. 1996; Fang and Hammes-Schiffer 1999; Granucci and Persico 2007; Shenvi et al. 2011a, b; Shenvi and Yang 2012; Subotnik and Shenvi 2011a, b; Agostini et al. 2016; Min et al. 2015, 2017), Ehrenfest dynamics and surface hopping (in their standard formulations) are not able to capture the decay of such quantity, observed here between 15 and 25 fs.

Additional information on the dynamics can be extracted from the analysis of the indicator of decoherence of Fig. 4. In fact, the pronounced double-peak structure suggests that two groups of trajectories funnel through the $S_1/S_2$ conical intersection at subsequent times. In order to interpret this observation, the indicator of decoherence has been decomposed in different contributions (represented by the colored curves in the lower panel of Fig. 4) arising from the different paths followed by the trajectories after crossing the conical intersection $S_1/S_2$. The structures identified at the end of the simulated trajectories are (i) a right-open ring structure (observed with probability 36%), (ii) a left-open ring structure (observed with probability 47%), (iii) a CC-extended bond structure (observed with probability 10%), and (iv) a closed-ring structure (observed with probability 7%). Structures (i) and (ii) are indeed equivalent; thus the difference in the probabilities can be probably cured by improving the statistics. These structures yield the ring-opening of oxirane via breaking of one of the two CO bonds. In structure (iii), the oxirane ring opens via the elongation of the CC bond. A few trajectories, identified as structure (iv), are not reactive, since the molecule more or less stays in its original configuration. If the indicator of decoherence is decomposed in contributions arising from trajectories ending in structures (i), (ii), or (ii), we observe that the first peak between 6 and 12 fs is produced by a first bundle of trajectories that leads to the breakage of the equivalent CO bonds. However, the $\eta_{12}(t)$ curves do not decay monotonically. Instead, the curves corresponding to the $C_1O$ and $C_2O$ groups both contribute to the second peak (between 12 fs and 17.5 fs). This feature indicates that the first group of trajectories is reached by a second group while funneling through the conical intersection. The main contribution to the second peak between 12 and 16 fs is given by trajectories yielding a final CC-extended bond structure. These trajectories clearly encounter the nonadiabatic region with some delay if compared to the sets of trajectories analyzed before. Here, the indicator of decoherence is clearly single peaked, suggesting that the corresponding trajectory bundle undergoes a transition through the $S_1/S_2$ conical intersection in a single step.

Observation of different final structures is related to the different paths undergone by the trajectories after crossing the $S_1/S_2$ region of nonadiabatic coupling. Two representative trajectories have been selected among the 100, one yielding a right-open ring structure (i) and one yielding a CC-extended bond configuration (iii). Clearly, the analysis presented for the right-open ring structure can be applied also to the equivalent left-open ring structure (ii). In Fig. 5 (upper panels), we show the electronic populations $|C_k^{(I)}(t)|^2$ as functions of time and for the selected trajectories, along with the energy profiles (lower panels) of the three adiabatic states considered here and the gauge-invariant (GI) part of the TDPES (the first two terms on the right-hand side of Eq. (14)). The TDPES provides information about the "active" electronic state: if one wants to connect the interpretation of the dynamics based on the exact factorization to the standard perspective in terms of wave packets evolving "on" different adiabatic surfaces, it is instructive to compare the TDPES with the BO PES, as done in Fig. 5 (lower panels).
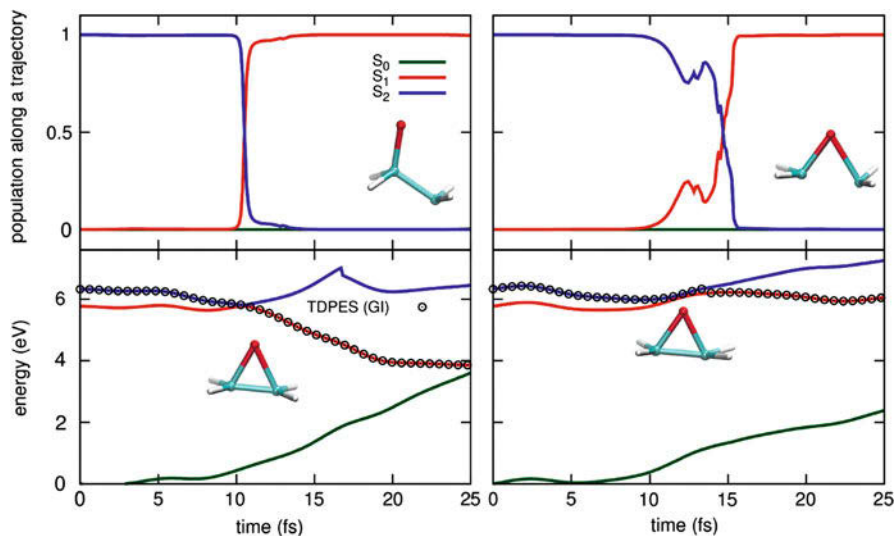
**Fig. 5** Upper panels: populations of the electronic states $S_0$, $S_1$, and $S_2$ as functions of time for two selected trajectories of type (i) (left) and of type (iii) (right). The color code is the same used in Fig. 4. Lower panels: energy profiles (in eV) along the selected trajectories, as in the upper panels. The zero is set to be the value of the energy of $S_0$ at time $t = 0$ fs. In the upper panels, a ball-stick representation of oxirane at the final time is shown, whereas in the lower panels, the configurations at the time of electronic population exchange are shown

The upper panels of Fig. 5 confirm that the coupling region is encountered by trajectories of type (iii) (right panels) at later times if compared to trajectories of type (i) or (ii) (left panels). Additionally, the $S_2/S_1$ population exchange is very sharp for type (i) and smooth for (iii). Observing the TDPES, we can argue that after about 5 fs, trajectories (i) encounter a steep $S_2$ potential, which directs them toward the conical intersection. Trajectories (iii) are trapped in a region of flat potential, which prevents them from a fast de-excitation to $S_1$. Subsequently, very different paths are undertaken, and thus different regions of the $S_1$ PES are explored. Toward the end of the simulated dynamics, only trajectories of type (i) are expected to relax to the ground state $S_0$, as confirmed by the closing of the energy gap between $S_0$ and $S_1$ at 25 fs (lower left panel of Fig. 5). At this time, the trajectory of type (iii) is evolving on a portion of the BO PES $S_1$ that is located at about 4 eV from the ground state.

## 4.4    Full and Ab Initio Multiple Spawning

### 4.4.1    Full Multiple Spawning

Full multiple spawning (FMS) (Martínez et al. 1996; Martínez and Levine 1997; Ben-Nun and Martínez 1998, 2002; Ben-Nun et al. 2000; Hack et al. 2001; Virshup

et al. 2008) proposes to expand the nuclear amplitudes in the Born-Huang expansion in a linear combination of frozen multidimensional Gaussian functions (Heller 1981). But these Gaussians functions do not form a static grid; on the contrary, they are evolving over time in both position and momentum space to better adapt to the evolution of the nuclear wavefunctions, forming a set of trajectory basis functions (TBFs):

$$
\Psi(\mathbf{r}, \mathbf{R}, t) = \sum_k \sum_I^{N_{TBFs,k}} C_I^{(k)}(t) \tilde{\chi}_I^{(k)} \left( \mathbf{R}; \overline{\mathbf{R}}_I^{(k)}(t), \overline{\mathbf{P}}_I^{(k)}(t), \overline{\gamma}_I^{(k)}(t), \boldsymbol{\alpha} \right) \varphi_{\mathbf{R}}^{(k)}(\mathbf{r}),
$$

(80)

where $C_I^{(k)}(t)$ is the complex coefficient for the TBF $I$ evolving on electronic state $(k)$ (used here as a label) and $\tilde{\chi}_I^{(k)} \left( \mathbf{R}; \overline{\mathbf{R}}_I^{(k)}(t), \overline{\mathbf{P}}_I^{(k)}(t), \overline{\gamma}_I^{(k)}(t), \boldsymbol{\alpha} \right)$ is the traveling multidimensional Gaussian $I$ on state $(k)$ with mean position $\overline{\mathbf{R}}_I^{(k)}(t)$, momenta $\overline{\mathbf{P}}_I^{(k)}(t)$, phase $\overline{\gamma}_I^{(k)}(t)$, and frozen width $\boldsymbol{\alpha}$. In FMS, the TBFs follow classical trajectories, i.e., $\overline{\mathbf{R}}(t)$ and $\overline{\mathbf{P}}(t)$ are propagated according to Hamilton's equation of motion (and the phase is integrated semi-classically). We note that this classical propagation of the TBFs does not imply that the method is semiclassical in itself, as the TBFs are only a support for the propagation of the nuclear wavefunctions. Indeed, in the limit of a large number of TBFs ($N_{TBFs}$), the FMS expansion would be exact. In fact, in the limit of an infinite number of Gaussian functions, their dynamics is redundant, and we have a (infinitely fine) grid. We note that other methods were proposed where the TBFs follow Ehrenfest trajectories (Shalashilin 2009, 2010; Saita and Shalashilin 2012; Makhov et al. 2017) (multiconfiguration Ehrenfest, MCE) or quantum trajectories (Worth et al. 2004, 2008; Lasorne et al. 2006, 2007; Mendive-Tapia et al. 2012; Richings et al. 2015) (variational Multiconfiguration Gaussian, vMCG), as well as mixed strategies (Makhov et al. 2014; Meek and Levine 2016; Izmaylov and Joubert-Doriol 2017; Joubert-Doriol et al. 2017).

One can express the time-dependent Schrödinger equation (1) in the basis of TBFs by inserting Eq. (80) in the former, left multiplying by

$$
\left[ \tilde{\chi}_J^{(l)} \left( \mathbf{R}; \overline{\mathbf{R}}_J^{(l)}(t), \overline{\mathbf{P}}_J^{(l)}(t), \overline{\gamma}_J^{(l)}(t), \boldsymbol{\alpha} \right) \varphi_{\mathbf{R}}^{(l)}(\mathbf{r}) \right]^*
$$

and integrating over both nuclear and electronic coordinates, leading, in atomic units, to (Ben-Nun and Martínez 2002):

$$
\frac{d}{dt} \mathbf{C}^l(t) = -i(\mathbf{S}_{ll}^{-1}) \left[ \left[ \mathbf{H}_{ll} - i\dot{\mathbf{S}}_{ll} \right] \mathbf{C}^l + \sum_{k \neq l} \mathbf{H}_{lk} \mathbf{C}^k \right].
$$

(81)

for each electronic state $l$ considered. The nonorthonormality of the Gaussian basis results in overlap matrices $\mathbf{S}_{ll}$ and $\dot{\mathbf{S}}_{ll}$, with elements $(\mathbf{S}_{ll})_{JI} = \langle \tilde{\chi}_J^{(l)} | \tilde{\chi}_I^{(l)} \rangle_{\mathbf{R}}$ and $(\dot{\mathbf{S}}_{ll})_{JI} = \langle \tilde{\chi}_J^{(l)} | \frac{\partial}{\partial t} \tilde{\chi}_I^{(l)} \rangle_{\mathbf{R}}$. We note that in Eq. (81) $\mathbf{S}_{lk} = \dot{\mathbf{S}}_{lk} = 0, \forall l \neq k$, due to the orthonormality of the electronic basis.

As mentioned earlier, the trajectories in FSSH are uncoupled. This is *not* the case in FMS, and TBFs are coupled thanks to the Hamiltonian matrix $\mathbf{H}$ in Eq. (81). Let us consider the Hamiltonian matrix element between two TBFs $J$ and $I$ evolving in adiabatic electronic states:

$$
\begin{aligned}
H_{kl}^{IJ} =& \langle \tilde{\chi}_I^{(k)} | \hat{T}_{\text{nuc}} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}} \delta_{kl} + \langle \tilde{\chi}_I^{(k)} | \varepsilon_{BO}^{(k)} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}} \delta_{kl} \\
& - \langle \tilde{\chi}_I^{(k)} | \sum_{\rho=1}^{3N} \frac{1}{M_\rho} \langle \varphi_{\mathbf{R}}^{(k)} | \frac{\partial}{\partial R_\rho} | \varphi_{\mathbf{R}}^{(l)} \rangle_{\mathbf{r}} \frac{\partial}{\partial R_\rho} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}} \\
& - \langle \tilde{\chi}_I^{(k)} | \sum_{\rho=1}^{3N} \frac{1}{2M_\rho} \langle \varphi_{\mathbf{R}}^{(k)} | \frac{\partial^2}{\partial R_\rho^2} | \varphi_{\mathbf{R}}^{(l)} \rangle_{\mathbf{r}} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}}
\end{aligned}
\tag{82}
$$

If the two TBFs are in the same electronic state $k$, they will be coupled *via* the first two terms on the r.h.s of Eq. (82): the nuclear kinetic energy operator and the (adiabatic) electronic energy. If the two TBFs belong to different electronic states $k$ and $l$, then a term containing the nonadiabatic coupling vectors (third term on the r.h.s) and the second-order nonadiabatic couplings (fourth term on the r.h.s of Eq. (82)) will ensure their coupling. We note that the diagonal second-order nonadiabatic couplings (also known as "diagonal Born-Oppenheimer correction") will contribute an additional intrastate coupling. FMS can easily incorporate additional source of couplings between TBFs like spin-orbit coupling (Curchod et al. 2016b), tunneling effects (Ben-Nun and Martínez 2000), or an external electromagnetic field (Mignolet et al. 2016), for example.

### 4.4.2 Spawning Algorithm

Up to this point, we have discussed the formal equations of FMS when a large number of TBFs is considered. However, FMS, as its name indicates, proposes to replace the large number of trajectory basis functions by an algorithm, coined the *spawning algorithm*, that will adapt the size of the basis set dynamically during the simulation. In other words, the spawning algorithm proposes the following alteration: $N_{TBFs} \rightarrow N_{TBFs}(t)$. In short, a TBF is evolving on a given PES, and as soon as a sizable coupling with a different electronic state is recorded, a spawning mode is triggered and will determine if, where, and when a new TBF should be created ("spawned") on the coupled state to maximize the coupling with the existing TBF and, therefore, the description of nonadiabatic transitions. Different versions of the spawning algorithm have been proposed, and the interested reader is encouraged to read Ben-Nun and Martínez (2002) and Yang et al. (2009) for more information. It is perhaps important to note at this stage that the time dependence of the number

of TBFs, $N_{TBFs}(t)$, implies that the size of matrices and vectors in Eq. (81) will vary during the dynamics.

### 4.4.3 Ab Initio Multiple Spawning

FMS is in principle exact. However, the coupling between TBFs given in Eq. (82) implies integration, and therefore knowledge, of electronic structure properties like PESs or nonadiabatic couplings over the full nuclear configuration space. Approximations are, therefore, needed to treat molecular systems in their full dimensionality. Let us start by performing a Taylor expansion of the electronic quantity of interest (e.g., here the electronic energy) around the centroid position of two TBFs $J$ and $I$ evolving in electronic state $k$: $\overline{\mathbf{R}}_{JI}^{(kk)} = \frac{\overline{\mathbf{R}}_J^{(k)} + \overline{\mathbf{R}}_I^{(k)}}{2}$:

$$\varepsilon_{BO}^{(k)}(\mathbf{R}) = \varepsilon_{BO}^{(k)}\left(\overline{\mathbf{R}}_{JI}^{(kk)}\right) + \sum_{\rho}^{3N} \left(R_\rho - \overline{R}_{\rho,JI}^{(kk)}\right) \frac{\partial \varepsilon_{BO}^{(k)}(\mathbf{R})}{\partial R_\rho}\Big|_{R_\rho = \overline{R}_{\rho,JI}^{(kk)}}$$

$$+ \frac{1}{2} \sum_{\rho\rho'}^{3N} \left(R_\rho - \overline{R}_{\rho,JI}^{(kk)}\right) \frac{\partial^2 \varepsilon_{BO}^{(k)}(\mathbf{R})}{\partial R_\rho \partial R_{\rho'}}\Big|_{R_\rho = \overline{R}_{\rho,JI}^{(kk)}, R_{\rho'} = \overline{R}_{\rho',JI}^{(kk)}} \left(R_{\rho'} - \overline{R}_{\rho',JI}^{(kk)}\right) + \dots$$

Owing to the locality of Gaussian functions, one will consider that, to a good approximation, only the term of order zero can be retained; in other words $\varepsilon_{BO}^{(k)}(\mathbf{R}) \approx \varepsilon_{BO}^{(k)}\left(\overline{\mathbf{R}}_{JI}^{(kk)}\right)$ (Ben-Nun and Martínez 2002). This approximation, called the saddle-point approximation of order zero (SPA0), strongly simplifies the coupling between TBFs as the Hamiltonian matrix elements become

$$H_{kl}^{IJ} = \langle \tilde{\chi}_I^{(k)} | \hat{T}_{nuc} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}} \delta_{kl} + \varepsilon_{BO}^{(k)}\left(\overline{\mathbf{R}}_{IJ}^{(kk)}\right) \langle \tilde{\chi}_I^{(k)} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}} \delta_{kl}$$

$$- \sum_{\rho=1}^{3N} \frac{1}{M_\rho} \langle \varphi_{\mathbf{R}}^{(k)} | \frac{\partial}{\partial R_\rho} | \varphi_{\mathbf{R}}^{(l)} \rangle_{\mathbf{r}}\Big|_{R_\rho = \overline{R}_{\rho,IJ}^{(kl)}} \langle \tilde{\chi}_I^{(k)} | \frac{\partial}{\partial R_\rho} | \tilde{\chi}_J^{(l)} \rangle_{\mathbf{R}} \quad (83)$$

(note that we also dropped the terms depending on second-order couplings). Hence, calculating the coupling between TBF $J$ and $I$ only requires the extra calculation of electronic structure quantities at their mutual centroid position, which can easily be achieved in an on-the-fly dynamics scheme. The SPA0 allows to port the *in principle* exact framework of FMS to the nonadiabatic dynamics of molecules. The resulting nonadiabatic method within the SPA0 is often called ab initio multiple spawning (AIMS). (We note that an additional approximation is commonly employed within AIMS – the independent first generation approximation – that approximates the initial nuclear wave packet at time $t = 0$ by a set of independent parent TBFs, i.e., coupling between parent TBFs is neglected. Couplings between all the TBFs spawned by each parent TBF are preserved (Ben-Nun and Martínez 2002)). AIMS has been coupled with different level of electronic structure

like SA-CASSCF (Levine et al. 2008), MS-CASPT2 (Tao et al. 2009), FOMO-CASCI (Pijeau et al. 2017), or LR-TDDFT (Curchod et al. 2017).

### 4.4.4    Applications

AIMS has recently (Curchod et al. 2017) been interfaced with an implementation of LR-TDDFT accelerated by graphical processing units (GPUs) (Isborn et al. 2011), offering an important speedup for the calculation of electronic energies, analytical gradients, and nonadiabatic coupling terms. The combination of AIMS and GPU-accelerated LR-TDDFT was employed to shed light on the excited-state dynamics of 4-N,N'-dimethylaminobenzonitrile (DMABN). DMABN is a molecule known to exhibit dual fluorescence depending on its environment, and it was proposed that the nature of the two emitting states is correlated with a twist of the dimethylamino (DMA) group (Grabowski et al. 2003) (see molecular structure in the inset of Fig. 6). However, DMABN is photoexcited into its second excited state, $S_2$, and relaxes to the first excited state $S_1$ where emission will occur at a later time. One question is then: does the $S_2/S_1$ nonadiabatic dynamics imply a twist of the DMA group, which could potentially have an influence on the fate of the molecule at later time in the $S_1$ state? Early static calculations have predicted that the $S_2/S_1$ transfer should be fast and that a twist of the DMA was not required for the nonadiabatic transition to occur (Gómez et al. 2005). AIMS combined with GPU-accelerated LR-TDDFT confirmed that the $S_2$ population decays to $S_1$ in less than $< 50$ fs (blue line, left panel of Fig. 6). Also, the transfer of the nuclear wave packet to $S_1$ is not correlated
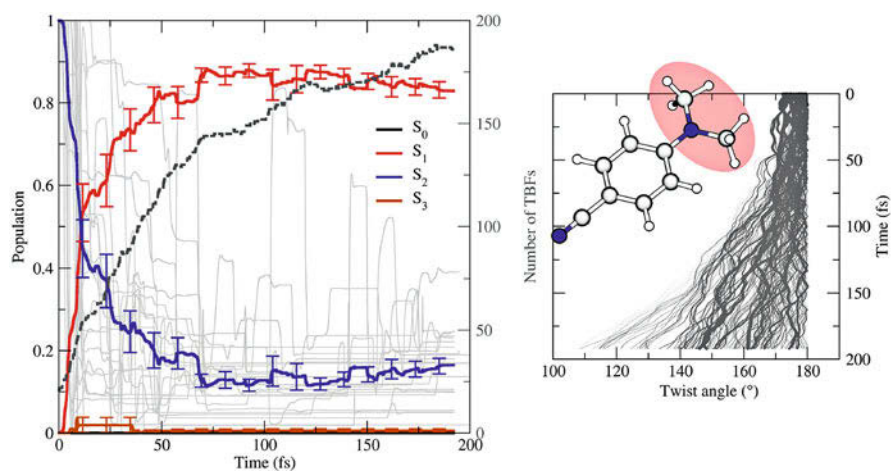


**Fig. 6** Left panel: Population traces (with standard errors) of the different electronic states considered in the AIMS/LR-TDDFT dynamics. The gray dashed line indicates the number of TBFs during the dynamics. Right panel: Twist angle of the DMA group for the entire swarm of TBFs. The line width is proportional to the TBF population. Inset: representation of the DMABN molecule, with the DMA group highlighted in red. (Adapted with permission from Curchod et al. 2016c. Copyright (2016) American Chemical Society)

with the torsion of the DMA group, as showed by the projection of the TBFs on the twist coordinate (right panel of Fig. 6). The DMA group in fact starts its rotation only after the molecule reached $S_1$ ($t > 50$ fs, see right panel of Fig. 6), and such a transfer implies a change in the electronic character of the wave packet from *charge transfer* in $S_2$ around the Franck-Condon region to *locally excited* after the nonadiabatic transfer to $S_1$. On a more technical note, the gray dashed line on the left panel of Fig. 6 indicates the number of TBFs during the AIMS dynamics. Starting from 21 parent TBFs, the spawning algorithm rapidly creates a large amount of child TBFs to ensure a good basis for the propagation in the different nonadiabatic regions encountered.

## 5    Conclusions

The aim of this chapter has been to provide a broad overview of quantum molecular dynamics methods to simulate nonadiabatic phenomena in isolated and condensed phase systems, adopting a quantum-classical perspective. The combination of classical and quantum-mechanical approaches is, in fact, capable to describe, accurately and efficiently, dynamical processes involving electronic excited states. The assumption that a good description of the nuclear dynamics can be achieved based on classical mechanics is based on the fact that at the molecular scale, atoms (usually heavier than a proton) are associated to a relatively short de Broglie wavelength, e.g., shorter than the typical scale of variation of the potential. Electrons, on the other hand, require a purely quantum-mechanical description, which in this work has been achieved using the framework of time-dependent density functional theory. The challenges faced in the development of theoretical and numerical approaches for nonadiabatic dynamics are copious, justifying the rise over the years of a multitude of strategies to tackle these types of problems, some of which have been described in this review chapter.

TDDFT has emerged as a very powerful method to describe electronic excited-state dynamics, both for molecular systems and in condensed phase. TDDFT describes the evolution of the electronic density in a time-dependent external potential, and within the linear-response regime, LR-TDDFT allows the calculation of excited-state properties, e.g., energies, forces, and nonadiabatic couplings at a modest computational cost. The straightforward combination of the fundamental TDDFT theorem (leading to the time-dependent Kohn-Sham equations) with the classical motion of the nuclei gives rise to the conceptually simple Ehrenfest dynamics scheme. We described some of its applications to molecular processes involving explicit time-dependent electronic wave packet dynamics, pointing out its limitations related to the mean-field character of the underlying approximations.

The strong interplay of electronic and nuclear motion is, however, highly nontrivial, and effects beyond the mean-field approximations are difficult to capture with trajectory-based approaches. In particular, the accurate description of the electronic "nonadiabatic effects" on the nuclear dynamics requires the derivation of a suitable theoretical framework that enables the separation of electronic and nuclear motions.

Building on the TDDFT or LR-TDDFT description of the electronic dynamics within either the Born-Huang or the exact factorization framework, we have then reviewed three quantum-classical nonadiabatic simulation techniques that extend beyond the mean-field Ehrenfest approach, namely, surface hopping, coupled-trajectory mixed quantum-classical dynamics, and multiple spawning, along with some of their recent applications. The common denominator of all these techniques is that they combine "on-the-fly" trajectory-based dynamics with the computational advantages of TDDFT.

In the past years, large progresses have thus been accomplished in mixed quantum-classical nonadiabatic dynamics, and we hope that this chapter will help newcomers to engage in this exciting field of research. We believe that TDDFT-based nonadiabatic dynamics can become the method of choice for treating photochemical and photophysical processes of complex molecular systems in the gas and condensed phases. While it is hard to predict the outcome of current and future developments, the improvement of the exchange-correlation functionals, for instance, going beyond the adiabatic approximation, is one of the key issues that should be addressed, aiming to get more reliable excited-state energies, nonadiabatic couplings, nuclear forces, and electronic dissipation effects.Finally, the development of coupling schemes to describe the interaction with electromagnetic fields described quantum mechanically is underway (Flick et al. 2017a, b; Dimitrov et al. 2017) and will open new research areas in the field of cavity quantum electrodynamics.

# References

Abedi A, Maitra NT, Gross EKU (2010) Exact factorization of the time-dependent electron-nuclear wave function. Phys Rev Lett 105(12):123002

Abedi A, Maitra NT, Gross EKU (2012) Correlated electron-nuclear dynamics: exact factorization of the molecular wave-function. J Chem Phys 137:22A530

Abedi A, Agostini F, Suzuki Y, Gross EKU (2013a) Dynamical steps that bridge piecewise adiabatic shapes in the exact time-dependent potential energy surface. Phys Rev Lett 110(26):263001

Abedi A, Maitra NT, Gross EKU (2013b) Reply to comment on "correlated electron-nuclear dynamics: exact factorization of the molecular wave-function". J Chem Phys 139(8):087102

Abedi A, Agostini F, Gross EKU (2014) Mixed quantum-classical dynamics from the exact decomposition of electron-nuclear motion. Europhys Lett 106(3):33001

Adamo C, Jacquemin D (2013) The calculations of excited-state properties with time-dependent density functional theory. Chem Soc Rev 42(3):845–856

Agostini F, Abedi A, Suzuki Y, Gross EKU (2013) Mixed quantum-classical dynamics on the exact time-dependent potential energy surfaces: a novel perspective on non-adiabatic processes. Mol Phys 111(22-23):3625

Agostini F, Abedi A, Gross EKU (2014) Classical nuclear motion coupled to electronic non-adiabatic transitions. J Chem Phys 141(21):214101

Agostini F, Abedi A, Suzuki Y, Min SK, Maitra NT, Gross EKU (2015a) The exact forces on classical nuclei in non-adiabatic charge transfer. J Chem Phys 142(8):084303

Agostini F, Min SK, Gross EKU (2015b) Semiclassical analysis of the electron-nuclear coupling in electronic non-adiabatic processes. Ann Phys 527(9–10):546–555

Agostini F, Min SK, Abedi A, Gross EKU (2016) Quantum-classical non-adiabatic dynamics: coupled- vs. independent-trajectory methods. J Chem Theory Comput 12(5):2127–2143

Agostini F, Tavernelli I, Ciccotti G (2018) Nuclear quantum effects in electronic (non)adiabatic dynamics. Eur Phys J B 91:139

Akimov AV, Prezhdo OV (2014) Advanced capabilities of the PYXAID program: integration schemes, decoherence effects, multiexcitonic states, and field-matter interaction. J Chem Theory Comput 10:789

Alonso JL, Clemente-Gallardo J, Echeniche-Robba P, Jover-Galtier JA (2013) Comment on "correlated electron-nuclear dynamics: exact factorization of the molecular wave-function". J Chem Phys 139:087101

Andrade X, Castro A, Zueco D, Alonso J, Echenique P, Falceto F, Rubio A (2009) Modified Ehrenfest formalism for efficient large-scale ab initio molecular dynamics. J Chem Theory Comput 5(4):728–742

Atkins AJ, González L (2017) Trajectory surface-hopping dynamics including intersystem crossing in [ru (bpy) 3] 2+. J Phys Chem Lett 8(16):3840–3845

Baer R (2002) Non-adiabatic couplings by time-dependent density functional theory. Chem Phys Lett 364:75–79

Baer M (2006) Beyond Born-Oppenheimer: electronic nonadiabatic coupling terms and conical intersections. Wiley, Hoboken

Barbatti M (2011) Nonadiabatic dynamics with trajectory surface hopping method. WIREs Comput Mol Sci 1:620–633

Ben-Nun M, Martínez TJ (1998) Nonadiabatic molecular dynamics: Validation of the multiple spawning method for a multidimensional problem. J Chem Phys 108:7244–7257

Ben-Nun M, Martínez TJ (2002) Ab initio quantum molecular dynamics. Adv Chem Phys 121:439–512

Ben-Nun M, Martínez TJ (2000) A multiple spawning approach to tunneling dynamics. J Chem Phys 112(14):6113–6121

Ben-Nun M, Quenneville J, Martínez TJ (2000) Ab initio multiple spawning: photochemistry from first principles quantum molecular dynamics. J Phys Chem A 104:5161–5175

Bittner ER, Rossky PJ (1995) Quantum decoherence in mixed quantum-classical systems: nonadiabatic processes. J Chem Phys 103:8130

Böckmann M, Doltsinis N, Marx D (2010) Unraveling a chemically enhanced photoswitch: bridged azobenzene. Angew Chemie Int Ed 49:3382

Bonella S, Coker DF (2005) LAND-map, a linearized approach to nonadiabatic dynamics using the mapping formalism. J Chem Phys 122:194102–194113

Burghardt I, Meyer HD, Cederbaum LS (1999) Approaches to the approximate treatment of complex molecular systems by the multiconfiguration time-dependent Hartree method. J Chem Phys 111:2927

Cannizzo A, van Mourik F, Gawelda W, Zgrablic G, Bressler C, Chergui M (2006) Broadband femtosecond fluorescence spectroscopy of $[Ru(bpy)_3]^{2+}$. Angew Chem Int Ed 45:3174–3176

Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. Phys Rev Lett 55:2471

Casida ME (1995) Time-dependent density-functional response theory for molecules. In: Chong DP (ed) Recent advances in density functional methods. World Scientific, Singapore, p 155

Casida ME (2009) Time-dependent density-functional theory for molecules and molecular solids. J Mol Struc (Theochem) 914(1–3):3–18

Casida M, Huix-Rotllant M (2012) Progress in time-dependent density-functional theory. Annu Rev Phys Chem 63(1):287–323. http://www.annualreviews.org/doi/pdf/10.1146/annurev-physchem-032511-143803

Casida ME, Gutierrez F, Guan J, Gadea FX, Salahub D, Daudey JP (2000) Charge-transfer correction for improved time-dependent local density approximation excited-state potential energy curves: analysis within the two-level model with illustration for $H_2O$ and LiH. J Chem Phys 113:7062

Castro A, Marques MAL, Rubio A (2004) Propagators for the time-dependent Kohn-Sham equations. J Chem Phys 121(8):3425–3433. https://doi.org/10.1063/1.1774980

Cave R, Zhang F, Maitra N, Burke K (2004) A dressed TDDFT treatment of the $2^1A_g$ states of butadiene and hexatriene. Chem Phys Lett 389(1):39–42

Chernyak V, Mukamel S (1996) Size-consistent quasiparticle representation of nonlinear optical susceptibilities in many-electron systems. J Chem Phys 104(2):444–459. https://doi.org/10.1063/1.470843

Chernyak V, Mukamel S (2000) Density-matrix representation of nonadiabatic couplings in time-dependent density functional (TDDFT) theories. J Chem Phys 112:3572–3579

Cordova F, Doriol LJ, Ipatov A, Casida ME, Filippi C, Vela A (2007) Troubleshooting time-dependent density-functional theory for photochemical applications: Oxirane. J Chem Phys 127:164111

Craig CF, Duncan WR, Prezhdo OV (2005) Trajectory surface hopping in the time-dependent Kohn-Sham approach for electron-nuclear dynamics. Phys Rev Lett 95:163001

Curchod BFE, Agostini F, Tavernelli I (2018) CT-MQC – a coupled-trajectory mixed quantum/classical method including nonadiabatic quantum coherence effects. Eur Phys J B 91:168

Curchod BFE, Agostini F (2017) On the dynamics through a conical intersection. J Phys Chem Lett 8:831

Curchod BFE, Tavernelli I (2013) On trajectory-based nonadiabatic dynamics: Bohmian dynamics versus trajectory surface hopping. J Chem Phys 138:184112

Curchod BFE, Tavernelli I, Rothlisberger U (2011) Trajectory-based solution of the nonadiabatic quantum dynamics equations: an on-the-fly approach for molecular dynamics simulations. Phys Chem Chem Phys 13:3231–3236

Curchod BFE, Rothlisberger U, Tavernelli I (2013) Trajectory-based nonadiabatic dynamics with time-dependent density functional theory. Chem Phys Chem 14(7):1314–1340

Curchod BFE, Agostini F, Gross EKU (2016a) An exact factorization perspective on quantum interferences in nonadiabatic dynamics. J Chem Phys 145:034103

Curchod BFE, Rauer C, Marquetand P, González L, Martínez T (2016b) Communication: Gaims–generalized ab initio multiple spawning for both internal conversion and intersystem crossing processes. J Chem Phys 144(10):101102

Curchod BFE , Sisto A, Martínez TJ (2016c) Ab initio multiple spawning photochemical dynamics of DMABN using GPUs. J Phys Chem A 121(1):265–276

Curchod BFE, Sisto A, Martínez TJ (2017) Ab initio multiple spawning photochemical dynamics of dmabn using GPUs. J Phys Chem A 121(1):265–276

Dancoff SM (1950) Non-adiabatic meson theory of nuclear forces. Phys Rev 78:382

Deglmann P, Furche F, Ahlrichs R (2002) An efficient implementation of second analytical derivatives for density functional methods. Chem Phys Lett 362(5–6):511–518. https://doi.org/10.1016/S0009-2614(02)01084-9, http://www.sciencedirect.com/science/article/pii/S0009261402010849

Dimitrov T, Flick J, Ruggenthaler M, Rubio A (2017) Exact functionals for correlated electron-photon systems. New J Phys 19:113036

Dobson JF, Bünner MJ, Gross EKU (1997) Time-dependent density functional theory beyond linear response: an exchange-correlation potential with memory. Phys Rev Lett 79(10):1905

Doltsinis NL, Marx D (2002) Nonadiabatic Car-Parrinello molecular dynamics. Phys Rev Lett 88:166402

Dreuw A, Head-Gordon M (2004) Failure of time-dependent density functional theory for long-range charge-transfer excited states: the zincbacteriochlorin-bacteriochlorin and bacteriochlorophyll-spheroidene complexes. J Am Chem Soc 126:4007–4016

Dreuw A, Head-Gordon M (2005) Single-reference ab initio methods for the calculation of excited states of large molecules. Chem Rev 105:4009

Dreuw A, Weisman J, Head-Gordon M (2003) Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. J Chem Phys 119:2943

Dunkel ER, Bonella S, Coker DF (2008) Iterative linearized approach to nonadiabatic dynamics. J Chem Phys 129:114106

Eich FG, Agostini F (2016) The adiabatic limit of the exact factorization of the electron-nuclear wave function. J Chem Phys 145:054110

Elliott P, Maitra NT (2012) Propagation of initially excited states in time-dependent density-functional theory. Phys Rev A 85:052510

Elliott P, Furche F, Burke K (2009) Excited states from time-dependent density functional theory. Rev Comput Chem 26:91

Elliott P, Goldson S, Canahui C, Maitra NT (2011) Perspectives on double-excitations in TDDFT. Chem Phys 391(1):110–119

Epstein S (1954) Note on perturbation theory. Am J Phys 22:613

Fang JY, Hammes-Schiffer S (1999) Improvement of the internal consistency in trajectory surface hopping. J Phys Chem A 103:9399–9407

Flick J, Appel H, Ruggenthaler M, Rubio A (2017a) Cavity Born-Oppenheimer approximation for correlated electron-nuclear-photon systems. J Chem Theory Comput 13:1616–1625

Flick J, Ruggenthaler M, Appel H, Rubio A (2017b) Atoms and molecules in cavities, from weak to strong coupling in quantum-electrodynamics (QED) chemistry. Proc Natl Acad Sci 114:3026–3034

Frenkel J (1934) Wave mechanics. Clarendon, Oxford

Furche F (2001) On the density matrix based approach to time-dependent density functional response theory. J Chem Phys 114:5982–5992

Gaigeot MP, Lopez-Tarifa P, Martin F, Alcami M, Vuilleumier R, Tavernelli I, Hervédu Penhoat MA, Politis MF (2010) Theoretical investigation of the ultrafast dissociation of ionised biomolecules immersed in water: direct and indirect effects. Mutat Res-Rev Mutat 704(1–3): 45–53. http://www.sciencedirect.com/science/article/pii/S1383574210000086

Gao X, Thiel W (2017) Non-hermitian surface hopping. Phys Rev E 95:013308

Garashchuk S, Rassolov VA (2003) Quantum dynamics with Bohmian trajectories: energy conserving approximation to the quantum potential. Chem Phys Lett 376:358

Gawelda W, Johnson M, de Groot FMF, Abela R, Bressler C, Chergui M (2006) Electronic and molecular structure of photoexcited $[Ru(II)(bpy)_3]^{2+}$ probed by picosecond x-ray absorption spectroscopy. J Am Chem Soc 128:5001–5009

Gómez I, Reguero M, Boggio-Pasqua M, Robb MA (2005) Intramolecular charge transfer in 4-aminobenzonitriles does not necessarily need the twist. J Am Chem Soc 127(19):7119–7129

Grabo T, Petersilka M, Gross EKU (2000) Molecular excitation energies from time-dependent density functional theory. J Mol Struc (Theochem) 501–502:353–367

Grabowski ZR, Rotkiewicz K, Rettig W (2003) Structural changes accompanying intramolecular electron transfer: focus on twisted intramolecular charge-transfer states and structures. Chem Rev 103(10):3899–4032

Granucci G, Persico M (2007) Critical appraisal of the fewest switches algorithm for surface hopping. J Chem Phys 126:134114

Gritsenko O, Baerends E (2004) Asymptotic correction of the exchange–correlation kernel of time-dependent density functional theory for long-range charge-transfer excitations. J Chem Phys 121:655

Gross EKU, Kohn W (1990) Time-dependent density-functional theory. Adv Quantum Chem 21: 255–291

Gross EKU, Kohn W (1985) Local density-functional theory of frequency-dependent linear response. Phys Rev Lett 55:2850–2852

Gross EKU, Ullrich CA, Gossmann UJ (1994) Density functional theory of time-dependent systems. In: Gross EKU, Dreizler RM (eds) Density functional theory. Plenum, New York, pp 149–171

Gross EKU, Dobson J, Petersilka M (1996) Density functional theory of time-dependent phenomena. In: Nalewajski RF (ed) Density functional theory II, topics in current chemistry, vol 181. Springer, Berlin, pp 81–172

Hack MD, Wensmann AM, Truhlar DG, Ben-Nun M, Martínez TJ (2001) Comparison of full multiple spawning, trajectory surface hopping, and converged quantum mechanics for electronically nonadiabatic dynamics. J Chem Phys 115:1172

Helgaker T, Jørgensen P (1989) Configuration-interaction energy derivatives in a fully variational formulation. Theor Chem Acc 75:111–127

Heller EJ (1981) Frozen gaussians: a very simple semiclassical approximation. J Chem Phys 75(6):2923–2931

Hellgren M, Gross EKU (2012) Discontinuities of the exchange-correlation kernel and charge-transfer excitations in time-dependent density-functional theory. Phys Rev A 85:022514

Hirata S, Head-Gordon M (1999) Time-dependent density functional theory within the Tamm-Dancoff approximation. Chem Phys Lett 314:291–299

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev B 136:B864

Hsu C, Hirata S, Head-Gordon M (2001) Excitation energies from time-dependent density functional theory for linear polyene oligomers: butadiene to decapentaene. J Phys Chem A 105(2):451–458

Hu C, Hirai H, Sugino O (2007) Nonadiabatic couplings from time-dependent density functional theory: formulation in the Casida formalism and practical scheme within modified linear response. J Chem Phys 127:064103

Hu C, Hirai H, Sugino O (2008) Nonadiabatic couplings from time-dependent density functional theory. II. Successes and challenges of the pseudopotential approximation. J Chem Phys 128:154111

Hu C, Sugino O, Hirai H, Tateyama Y (2010) Nonadiabatic couplings from the Kohn-Sham derivative matrix: formulation by time-dependent density-functional theory and evaluation in the pseudopotential framework. Phys Rev A 82(6):062508

Hu C, Komakura R, Li Z, Watanabe K (2012) TDDFT study on quantization behaviors of nonadiabatic couplings in polyatomic systems. Int J Quantum Chem 113:263–271

Huo P, Coker DF (2012) Consistent schemes for non-adiabatic dynamics derived from partial linearized density matrix propagation. J Chem Phys 137:22A535

Hutter J (2003) Excited state nuclear forces from the Tamm-Dancoff approximation to time-dependent density functional theory within the plane wave basis set framework. J Chem Phys 118:3928–3934

Iikura H, Tsuneda T, Yanai T, Hirao K (2001) A long-range correction scheme for generalized-gradient-approximation exchange functionals. J Chem Phys 115:3540

Isborn CM, Luehr N, Ufimtsev IS, Martínez TJ (2011) Excited-state electronic structure with configuration interaction singles and Tamm–Dancoff time-dependent density functional theory on graphical processing units. J Chem Theory Comput 7(6):1814

Izmaylov AF, Joubert-Doriol L (2017) Quantum nonadiabatic cloning of entangled coherent states. J Phys Chem Lett 8(8):1793–1797

Jaeger HM, Fischer S, Prezhdo OV (2012) Decoherence-induced surface hopping. J Chem Phys 137:22A545

Jamorski C, Casida ME, Salahub DR (1996) Dynamic polarizabilities and excitation spectra from a molecular implementation of time-dependent density-functional response theory: $N_2$ as a case study. J Chem Phys 104:5134

Jasper AW, Truhlar DG (2007) Electronic decoherence time for non-Born-Oppenheimer trajectories. J Chem Phys 127:194306

Jasper AW, Zhu C, Nangia S, Truhlar DG (2004) Introductory lecture: nonadiabatic effects in chemical dynamics. Faraday Discuss 127:1

Jasper AW, Nangia S, Zhu C, Truhlar DG (2006) Non-born-oppenheimer molecular dynamics. Acc Chem Res 39:101

Joubert-Doriol L, Sivasubramanium J, Ryabinkin IG, Izmaylov AF (2017) Topologically correct quantum nonadiabatic formalism for on-the-fly dynamics. J Phys Chem Lett 8(2):452–456

Kapral R (2006) Progress in the theory of mixed quantum-classical dynamics. Annu Rev Phys Chem 57(1):129–157

Kapral R, Ciccotti G (1999) Mixed quantum-classical dynamics. J Chem Phys 110(18):8919–8929. https://doi.org/10.1063/1.478811

Kleinman L, Bylander DM (1982) Efficacious form for model pseudopotentials. Phys Rev Lett 48:1425

Kurzweil Y, Baer R (2004) Time-dependent exchange-correlation current density functionals with memory. J Chem Phys 121(18):8731–8741. https://doi.org/10.1063/1.1802793

Lara-Astiaso M, Palacios A, Decleva P, Tavernelli I, Martín F (2017) Role of electron-nuclear coupled dynamics on charge migration induced by attosecond pulses in glycine. Chem Phys Lett 683:357

Lasorne B, Bearpark MJ, Robb MA, Worth GA (2006) Direct quantum dynamics using variational multi-configuration gaussian wavepackets. Chem Phys Lett 432(4):604–609

Lasorne B, Robb M, Worth G (2007) Direct quantum dynamics using variational multi-configuration gaussian wavepackets. implementation details and test case. Phys Chem Chem Phys 9(25):3210–3227

Laurent AD, Jacquemin D (2013) Td-dft benchmarks: a review. Int J Quant Chem 113(17):2019–2039

Lauvergnat D, Nauts A (2010) Torsional energy levels of nitric acid in reduced and full dimensionality with elvibrot and tnum. Phys Chem Chem Phys 12:8405

Lauvergnat D, Nauts A (2014) Quantum dynamics with sparse grids: a combination of Smolyak scheme and cubature. Application to methanol in full dimensionality. Spectrochim Acta Part A 119:18

Leininger T, Stoll H, Werner H, Savin A (1997) Combining long-range configuration interaction with short-range density functionals. Chem Phys Lett 275(3):151–160

Levine BG, Ko C, Quenneville J, Martinez TJ (2006) Conical intersections and double excitations in density functional theory. Mol Phys 104:1039

Levine BG, Coe JD, Virshup AM, Martinez TJ (2008) Implementation of ab initio multiple spawning in the molpro quantum chemistry package. Chem Phys 347(1):3–16

Li X, Tully JC, Schlegel HB, Frisch MJ (2005) Ab initio Ehrenfest dynamics. J Chem Phys 123(8):084106. https://doi.org/10.1063/1.2008258

Li Z, Liu W (2014) First-order nonadiabatic coupling matrix elements between excited states: a lagrangian formulation at the CIS, RPA, TD-HF, and TD-DFT levels. J Chemi Phys 141(1):014110

Li Z, Suo B, Liu W (2014) First order nonadiabatic coupling matrix elements between excited states: implementation and application at the TD-DFT and PP-TDA levels. J Chem Phys 141(24):244105

Liang W, Isborn CM, Lindsay A, Li X, Smith SM, Levis RJ (2010) Time-dependent density functional theory calculations of Ehrenfest dynamics of laser controlled dissociation of $NO^+$: Pulse length and sequential multiple single-photon processes. J Phys Chem A 114(21):6201–6206

Lopez-Tarifa P, Herve du Penhoat MA, Vuilleumier R, Gaigeot MP, Tavernelli I, Le Padellec A, Champeaux JP, Alcami M, Moretto-Capelle P, Martin F, Politis MF (2011) Ultrafast nonadiabatic fragmentation dynamics of doubly charged uracil in a gas phase. Phys Rev Lett 107:023202

Lopreore CL, Wyatt RE (2002) Electronic transitions with quantum trajectories. II. J Chem Phys 116(4):1228–1238

Maitra NT (2005) Undoing static correlation: long-range charge transfer in time-dependent density-functional theory. J Chem Phys 122:234104

Maitra NT, Wasserman A, Burke K (2003) What is time-dependent density-functional theory? successes and challenges. In: Gonis A, Kioussis N, Ciftan M (eds) Electron correlations and materials properties 2. Kluwer/Plenum, New York

Maitra NT, Zhang F, Cave RJ, Burke K (2004) Double excitations within time-dependent density functional theory linear response. J Chem Phys 120:5932

Makhov DV, Glover WJ, Martinez TJ, Shalashilin DV (2014) Ab initio multiple cloning algorithm for quantum nonadiabatic molecular dynamics. J Chem Phys 141(5):054110

Makhov D, Symonds C, Fernandez-Alberti S, Shalashilin D (2017) Ab initio quantum direct dynamics simulations of ultrafast photochemistry with multiconfigurational ehrenfest approach. Chem Phys 493:200–218

Marques MAL, Maitra NT, Nogueira FMDS, Gross EKU, Rubio A (2012) Fundamentals of time-dependent density functional theory, vol 837. Springer, Berlin/Heidelberg

Martínez TJ, Levine RD (1997) Non-adiabatic molecular dynamics: split-operator multiple spawning with applications to photodissociation. J Chem Soc Faraday Trans 93(5):941–947

Martínez TJ, Ben-Nun M, Levine RD (1996) Multi-electronic-state molecular dynamics: a wave function approach with applications. J Phys Chem 100(19):7884–7895

Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge

Meek GA, Levine BG (2016) The best of both reps—diabatized gaussians on adiabatic surfaces. J Chem Phys 145(18):184103

Mendive-Tapia D, Lasorne B, Worth GA, Robb MA, Bearpark MJ (2012) Towards converging non-adiabatic direct dynamics calculations using frozen-width variational gaussian product basis functions. J Chem Phys 137(22):22A548

Meyer HD, Worth GA (2003) Quantum molecular dynamics: propagating wavepackets and density operators using the multiconfiguration time-dependent hartree method. Theor Chim Acta 109:251

Meyer HD, Manthe U, Cederbaum LS (1990) The multi-configurational time-dependent hartree approach. Chem Phys Lett 165:73–78

Mignolet B, Curchod BFE, Martínez TJ (2016) Communication: Xfaims—external field ab initio multiple spawning for electron-nuclear dynamics triggered by short laser pulses. J Chem Phys 145(19):191104

Min SK, Agostini F, Gross EKU (2015) Coupled-trajectory quantum-classical approach to electronic decoherence in nonadiabatic processes. Phys Rev Lett 115(7):073001

Min SK, Agostini F, Tavernelli I, Gross EKU (2017) Ab initio nonadiabatic dynamics with coupled trajectories: a rigorous approach to quantum (de)coherence. J Phys Chem Lett 8:3048

Moss CL, Isborn CM, Li X (2009) Ehrenfest dynamics with a time-dependent density-functional-theory calculation of lifetimes and resonant widths of charge-transfer states of $Li^+$ near an aluminum cluster surface. Phys Rev A 80:024503. https://doi.org/10.1103/PhysRevA.80.024503

Nielsen S, Kapral R, Ciccotti G (2000) Non-adiabatic dynamics in mixed quantum-classical systems. J Stat Phys 101:225–242

Ou Q, Bellchambers GD, Furche F, Subotnik JE (2015) First-order derivative couplings between excited states from adiabatic TDDFT response theory. J Chem Phys 142(6):064114

Parker SM, Roy S, Furche F (2016) Unphysical divergences in response theory. J Chem Phys 145(13):134105

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865

Persico M, Granucci G (2014) An overview of nonadiabatic dynamics simulations methods, with focus on the direct approach versus the fitting of potential energy surfaces. Theor Chem Acc 133(9):1–28

Petersilka M, Gossmann UJ, Gross EKU (1996) Excitation energies from time-dependent density-functional theory. Phys Rev Lett 76:1212–1215

Pijeau S, Foster D, Hohenstein EG (2017) Excited-state dynamics of 2-(2'-hydroxyphenyl) benzothiazole: ultrafast proton transfer and internal conversion. J Phys Chem A 121:4595

Pulay P (1987) Analytical derivative methods in quantum chemistry. Adv Chem Phys 69:241–286

Rappoport D, Furche F (2005) Analytical time-dependent density functional derivative methods within the RI-*J* approximation, an approach to excited states of large molecules. J Chem Phys 122(6):064105. https://doi.org/10.1063/1.1844492

Rassolov VA, Garashchuk S (2005) Semiclassical nonadiabatic dynamics with quantum trajectories. Phys Rev A 71(3):032511

Requist R, Gross EKU (2016) Exact factorization-based density functional theory of electrons and nuclei. Phys Rev Lett 117:193001

Richings G, Polyak I, Spinlove K, Worth G, Burghardt I, Lasorne B (2015) Quantum dynamics simulations using gaussian wavepackets: the vMCG method. Int Rev Phys Chem 34(2): 269–308

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52:997–1000

Sadri K, Lauvergnat D, Gatti F, Meyer HD (2012) Numeric kinetic energy operators for molecules in polyspherical coordinates. J Chem Phys 136:234112

Sadri K, Lauvergnat D, Gatti F, Meyer HD (2014) Rovibrational spectroscopy using a kinetic energy operator in Eckart frame and the multi-configuration time-dependent hartree (MCTDH) approach. J Chem Phys 141:114101

Saita K, Shalashilin DV (2012) On-the-fly ab initio molecular dynamics with multiconfigurational ehrenfest method. J Chem Phys 137(22):22A506

Scherrer A, Agostini F, Sebastiani D, Gross EKU, Vuilleumier R (2015) Nuclear velocity perturbation theory for vibrational circular dichroism: an approach based on the exact factorization of the electron-nuclear wave function. J Chem Phys 143(7):074106

Scherrer A, Agostini F, Sebastiani D, Gross EKU, Vuilleumier R (2017) On the mass of atoms in molecules: beyond the born-oppenheimer approximation. Phys Rev X 7:031035

Schild A, Agostini F, Gross EKU (2016) Electronic flux density beyond the born-oppenheimer approximation. J Phys Chem A 120:3316

Schwartz BJ, Bittner ER, Prezhdo OV, Rossky PJ (1996) Quantum decoherence and the isotope effect in condensed phase nonadiabatic molecular dynamics simulations. J Chem Phys 104:5942

Send R, Furche F (2010) First-order nonadiabatic couplings from time-dependent hybrid density functional response theory: consistent formalism, implementation, and performance. J Chem Phys 132(4):044107. https://doi.org/10.1063/1.3292571

Shalashilin D (2009) Quantum mechanics with the basis set guided by ehrenfest trajectories: theory and application to spin-boson model. J Chem Phys 130:244101

Shalashilin DV (2010) Nonadiabatic dynamics with the help of multiconfigurational ehrenfest method: improved theory and fully quantum 24d simulation of pyrazine. J Chem Phys 132(24):244111

Shenvi N, Yang W (2012) Achieving partial decoherence in surface hopping through phase correction. J Chem Phys 137:22A528

Shenvi N, Subotnik JE, Yang W (2011a) Phase-corrected surface hopping: correcting the phase evolution of the electronic wavefunction. J Chem Phys 135:024101

Shenvi N, Subotnik JE, Yang W (2011b) Simultaneous-trajectory surface hopping: a parameter-free algorithm for implementing decoherence in nonadiabatic dynamics. J Chem Phys 134:144102

Stratmann RE, Scuseria GE, Frisch MJ (1998) An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules. J Chem Phys 109(19):8218–8224

Subotnik JE, Shenvi N (2011a) Decoherence and surface hopping: when can averaging over initial conditions help capture the effects of wave packet separation? J Chem Phys 134:244114

Subotnik JE, Shenvi N (2011b) A new approach to decoherence and momentum rescaling in the surface hopping algorithm. J Chem Phys 134:024105

Subotnik JE, Ouyang W, Landry BR (2013) Can we derive Tully's surface-hopping algorithm from the semiclassical quantum Liouville equation? Almost, but only with decoherence. J Chem Phys 139:214107

Suzuki Y, Watanabe K (2016) Bohmian mechanics in the exact factorization of electron-nuclear wave functions. Phys Rev A 94:032517

Suzuki Y, Abedi A, Maitra NT, Gross EKU (2015) Laser-induced electron localization in $H_2^+$: Mixed quantum-classical dynamics based on the exact time-dependent potential energy surface. Phys Chem Chem Phys 17:29271–29280

Tamm I (1945) Relativistic interaction of elementary particles J Phys 9:449

Tao H, Levine BG, Martínez TJ (2009) Ab initio multiple spawning dynamics using multi-state second-order perturbation theory. J Chem Phys A 113(49):13656–13662

Tapavicza E, Tavernelli I, Rothlisberger U (2007) Trajectory surface hopping within linear response time-dependent density-functional theory. Phys Rev Lett 98:023001

Tapavicza E, Tavernelli I, Rothlisberger U, Filippi C, Casida ME (2008) Mixed time-dependent density-functional theory/classical trajectory surface hopping study of oxirane photochemistry. J Chem Phys 129:124108

Tavernelli I (2006) Electronic density response of liquid water using time-dependent density functional theory. Phys Rev B 73:094204

Tavernelli I (2013) Ab initio–driven trajectory-based nuclear quantum dynamics in phase space. Phys Rev A 87(4):042501

Tavernelli I (2015) Nonadiabatic molecular dynamics simulations: synergies between theory and experiments. Acc Chem Res 48(3):792–800

Tavernelli I, Röhrig U, Rothlisberger U (2005) Molecular dynamics in electronically excited states using time-dependent density functional theory. Mol Phys 103(6–8):963–981

Tavernelli I, Curchod BFE, Rothlisberger U (2009a) On nonadiabatic coupling vectors in time-dependent density functional theory. J Chem Phys 131:196101

Tavernelli I, Tapavicza E, Rothlisberger U (2009b) Nonadiabatic coupling vectors within linear response time-dependent density functional theory. J Chem Phys 130:124107

Tavernelli I, Curchod BFE, Laktionov A, Rothlisberger U (2010) Nonadiabatic coupling vectors for excited states within time-dependent density functional theory and beyond. J Chem Phys 133:194104–194110

Tavernelli I, Curchod BFE, Rothlisberger U (2011) Nonadiabatic molecular dynamics with solvent effects: a LR-TDDFT QM/MM study of ruthenium (II) tris (bipyridine) in water. Chem Phys 391:101

Tozer D (2003) Relationship between long-range charge-transfer excitation energy error and integer discontinuity in Kohn–Sham theory. J Chem Phys 119:12697

Tozer DJ, Handy NC (2000) On the determination of excitation energies using density functional theory. Phys Chem Chem Phys 2(10):2117–2121

Tully JC (1990) Molecular dynamics with electronic transitions. J Chem Phys 93:1061

Tully JC (1998) Mixed quantum-classical dynamics. Faraday Discuss 110:407

Ullrich CA (2012) Time-dependent density-functional theory. Oxford, Oxford University Press

Ullrich CA, Tokatly IV (2006) Nonadiabatic electron dynamics in time-dependent density-functional theory. Phys Rev B 73:235102. http://link.aps.org/doi/10.1103/PhysRevB.73.235102

Vacher M, Bearpark MJ, Robb MA, Malhado JP (2017) Electron dynamics upon ionization of polyatomic molecules: Coupling to quantum nuclear motion and decoherence. Phys Rev Lett 118(8):083001

van Leeuwen R (1998) Causality and symmetry in time-dependent density-functional theory. Phys Rev Lett 80(6):1280–1283

van Leeuwen R (1999) Mapping from densities to potentials in time-dependent density-functional theory. Phys Rev Lett 82(19):3863–3866

Van Vleck JH (1928) The correspondence principle in the statistical interpretation of quantum mechanics. Proc Nat Aca Sci USA 14(2):178

Vignale G (2008) Real-time resolution of the causality paradox of time-dependent density-functional theory. Phys Rev A 77(6):062511

Virshup AM, Punwong C, Pogorelov TV, Lindquist BA, Ko C, Martínez TJ (2008) Photodynamics in complex environments: ab initio multiple spawning quantum mechanical/molecular mechanical dynamics. J Phys Chem B 113(11):3280–3291

Wang F, Ziegler T (2005) A simplified relativistic time-dependent density-functional theory formalism for the calculations of excitation energies including spin-orbit coupling effect. J Chem Phys 123(15):154102

Wang H, Thoss M (2003) Multilayer formulation of the multiconfiguration time-dependent Hartree theory. J Chem Phys 119:1289

Wiggins P, Williams JAG, Tozer DJ (2009) Excited state surfaces in density functional theory: a new twist on an old problem. J Chem Phys 131(9):091101

Wijewardane HO, Ullrich CA (2008) Real-time electron dynamics with exact-exchange time-dependent density-functional theory. Phys Rev Lett 100:056404. http://link.aps.org/doi/10.1103/PhysRevLett.100.056404

Worth G, Robb M, Burghardt I (2004) A novel algorithm for non-adiabatic direct dynamics using variational Gaussian wavepackets. Faraday Discuss 127:307–323

Worth GA, Robb MA, Lasorne B (2008) Solving the time-dependent Schrödinger equation for nuclear motion in one step: direct dynamics of non-adiabatic systems. Mol Phys 106(16–18):2077–2091

Wyatt RE, Lopreore CL, Parlant G (2001) Electronic transitions with quantum trajectories. J Chem Phys 114(12):5113–5116

Yagi K, Takatsuka K (2005) Nonadiabatic chemical dynamics in an intense laser field: electronic wave packet coupled with classical nuclear motions. J Chem Phys 123(22):224103

Yanai T, Tew D, Handy N (2004) A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). Chem Phys Lett 393:51

Yang S, Coe JD, Kaduk B, Martínez TJ (2009) An "optimal" spawning algorithm for adaptive basis set expansion in nonadiabatic dynamics. J Chem Phys 130(13):04B606

Zangwill A, Soven P (1980) Density-functional approach to local-field effects in finite systems: Photoabsorption in the rare gases. Phys Rev A 21(5):1561

# Ab Initio Electronic Structure Calculations by Auxiliary-Field Quantum Monte Carlo

**7**

Shiwei Zhang

## Contents

## Abstract

The auxiliary-field quantum Monte Carlo (AFQMC) method provides a computational framework for solving the time-independent Schrödinger equation in atoms, molecules, solids, and a variety of model systems by stochastic sampling. We introduce the theory and formalism behind this framework, briefly discuss the key technical steps that turn it into an effective and practical computational method, present several illustrative results, and conclude with comments on the prospects of ab initio computation by this framework.

S. Zhang (✉)
Center for Computational Quantum Physics, Flatiron Institute, New York, NY, USA

Department of Physics, College of William and Mary, Williamsburg, VA, USA
e-mail: szhang@flatironinstitute.org

# 1    Introduction

Predicting materials properties requires robust and reliable calculations at the most fundamental level. Often the effects being studied or designed originate from electron correlations, and small errors in their treatment can result in crucial and qualitative differences in the properties. The accurate treatment of interacting quantum systems is one of the grand challenges in modern science. In condensed phase materials, the challenge is increased by the need to account for the interplay between the electrons and the chemical and structural environment. Progress in addressing this challenge will be fundamental to achieving the materials genome initiative.

Explicit solution of the many-body Schrödinger equation leads to rapidly growing computational cost as a function of system size (see, e.g., Szabo and Ostlund 1989). To circumvent the problem, most computational quantum mechanical studies of large, realistic systems rely on simpler independent-particle approaches based on density-functional theory (DFT) (see, e.g., Kohn 1999; Martin 2004), using an approximate energy functional to include many-body effects. These replace the electron-electron interaction by an effective potential, thereby reducing the problem to a set of one-electron equations. Methods based on DFT and through its Car-Parrinello molecular dynamics implementation (Car and Parrinello 1985) have been extremely effective in complex molecules and solids (Kohn 1999). Such approaches are the standard in electronic structure, widely applied in condensed matter, quantum chemistry, and materials science.

Despite the tremendous successes of DFT, the treatment of electronic correlation is approximate. For strongly correlated systems (e.g., high-temperature superconductors, heavy-fermion metals, magnetic materials, optical lattices), where correlation effects from particle interaction crucially modify materials properties, the approximation can lead to qualitatively incorrect results. Even in moderately correlated systems when the method is qualitatively correct, the results are sometimes not sufficiently accurate. For example, in ferroelectric materials the usually acceptable 1% errors in DFT predictions of the equilibrium lattice constant can lead to almost full suppression of the ferroelectric order.

The development of alternatives to independent-particle theories is therefore of paramount fundamental and practical significance. To accurately capture the quantum many-body effects, the size of the Hilbert space involved often grows exponentially. Simulation methods utilizing Monte Carlo (MC) sampling (Kalos et al. 1974; Foulkes et al. 2001; Ceperley 1995; Blankenbecler et al. 1981; Sugiyama and Koonin 1986; Zhang and Krakauer 2003) are, in principle, both non-perturbative and well-equipped to handle details and complexities in the external environment. They are a unique combination of accuracy, general applicability, favorable scaling (low-power) for computational cost with physical system size, and scalability on parallel computing platforms (Esler et al. 2008).

For fermion systems, however, a so-called "sign" problem (Schmidt and Kalos 1984; Loh et al. 1990; Zhang 1999a) arises in varying forms in these MC simulation methods. The Pauli exclusion principle requires that the states be antisymmetric

under interchange of two particles. As a consequence, negative signs appear, which cause cancelations among contributions of the MC samples of the wave function or density matrix. In some formalism, as discussed below, a phase appears which leads to a continuous degeneracy and more severe cancelations. As the temperature is lowered or the system size is increased, such cancelation becomes more and more complete. The net signal thus decays *exponentially* versus noise. The algebraic scaling is then lost, and the method breaks down. Clearly the impact of this problem on the study of correlated electron systems is extremely severe.

In this chapter, we discuss the auxiliary-field quantum Monte Carlo (AFQMC) method for many-body computations in real materials. We cast the MC random walks in a space of over-complete Slater determinants, which significantly reduce the severity of the sign problem. In this space we formulate constraints on the random walk paths which lead to better approximations that are less sensitive to the details of the constraint. We then develop internal checks and constraint release methods to systematically improve the approach. These methods have come under the name of constrained path Monte Carlo (CPMC) (Zhang et al. 1997) for systems where there is a sign problem (e.g., Hubbard-like models where the auxiliary fields are real due to the short-ranged interactions). For electronic systems where there is a phase problem (as the Coulomb interaction leads to complex fields), the methods have been referred to as phaseless AFQMC (Zhang and Krakauer 2003; Al-Saidi et al. 2006; Motta and Zhang 2018). Here we will refer to the method as AFQMC; when necessary to emphasize the constrained-path (CP) approximation to distinguish the method from unconstrained free-projection, we will refer to it as CP-AFQMC.

## 2 Formalism

The Hamiltonian for any many-fermion system with two-body interactions (e.g., the electronic Hamiltonian under the Born-Oppenheimer approximation) can be written as

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = -\frac{\hbar^2}{2m}\sum_{m=1}^{M}\nabla_m^2 + \sum_{m=1}^{M}V_{\text{ext}}(\mathbf{r}_m) + \sum_{m<n}^{M}V_{\text{int}}(\mathbf{r}_m - \mathbf{r}_n), \qquad (1)$$

where $\mathbf{r}_m$ is the real-space coordinate of the $m$-th fermion. The one-body part of the Hamiltonian, $\hat{H}_1$, consists of the kinetic energy of the electrons and the effect of the ionic (and any other external) potentials. (We have represented the external potential as local, although this does not have to be the case. For example, in plane-wave calculations, we will use a norm-conserving pseudopotential, which will lead to a nonlocal function $V_{\text{ext}}$.) The two-body part of the Hamiltonian, $\hat{H}_2$, contains the electron-electron interaction terms. The total number of fermions, $M$, will be fixed in the calculations we discuss. For simplicity, we have suppressed spin-index, but the spin will be made explicit when necessary. In that case, $M_\sigma$ is the number

of electrons with spin $\sigma$ ($\sigma = \uparrow$ or $\downarrow$). We assume that the interaction is spin-independent, so the total $S_z$, defined by $(M_\uparrow - M_\downarrow)$, is fixed in the calculation, although it will be straightforward to generalize our discussions to treat other cases, for example, when there is spin-orbit coupling (SOC) (Rosenberg et al. 2017).

With any chosen one-particle basis, the Hamiltonian can be written in second quantization in the general form

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = \sum_{i,j}^{N} T_{ij} c_i^\dagger c_j + \frac{1}{2} \sum_{i,j,k,l}^{N} V_{ijkl} c_i^\dagger c_j^\dagger c_k c_l , \tag{2}$$

where the one-particle basis, $\{|\chi_i\rangle\}$ with $i = 1, 2, \cdots, N$, can be lattice sites (Hubbard model), plane waves (as in solid state calculations) (Suewattana et al. 2007), Gaussians (as in quantum chemistry) (Al-Saidi et al. 2006; Purwanto et al. 2011), etc. The operators $c_i^\dagger$ and $c_i$ are creation and annihilation operators on $|\chi_i\rangle$, satisfying standard fermion commutation relations. The one-body matrix elements, $T_{ij}$, contain the effect of both the kinetic energy and external potential, while the two-body matrix elements, $V_{ijkl}$, are from the interaction. The matrix elements are expressed in terms of the basis functions, for example,

$$V_{ijkl} = \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_i^*(\mathbf{r}_1) \chi_j^*(\mathbf{r}_2) V_{int}(\mathbf{r}_1 - \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_1) . \tag{3}$$

In quantum chemistry calculations, these are readily evaluated with standard Gaussian basis sets. In solid state calculations with plane waves, the kinetic and electron-electron interaction terms have simple analytic expressions, while the electron-ion potential leads to terms which are provided by the pseudopotential generation. We will assume that all matrix elements in Eq. (2) have been evaluated and are known as we begin our many-body calculations.

## 2.1   Non-orthogonal Slater Determinant Space

The AFQMC method seeks to obtain the ground state of the Hamiltonian in Eq. (2), representing it stochastically in the form

$$|\Psi_0\rangle = \sum_\phi \alpha_\phi |\phi\rangle , \tag{4}$$

where $|\phi\rangle$ is a Slater determinant:

$$|\phi\rangle \equiv \hat{\varphi}_1^\dagger \hat{\varphi}_2^\dagger \cdots \hat{\varphi}_M^\dagger |0\rangle . \tag{5}$$

In Eq. (5), the operator $\hat{\varphi}_m^\dagger \equiv \sum_i c_i^\dagger \varphi_{i,m}$, with $m$ taking an integer value among $1, 2, \cdots, M$, creates an electron in a single-particle orbital $\varphi_m$: $\hat{\varphi}_m^\dagger |0\rangle = \sum_i \varphi_{i,m} |\chi_i\rangle$. The content of the orbital can thus be conveniently expressed as an

$N$-dimensional vector $\{\varphi_{1,m}, \varphi_{2,m}, \cdots, \varphi_{N,m}\}$. The Slater determinant $|\phi\rangle$ in Eq. (5) can then be expressed as an $N \times M$ matrix:

$$\Phi \equiv \begin{pmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,M} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,M} \\ \vdots & \vdots & & \vdots \\ \varphi_{N,1} & \varphi_{N,2} & \cdots & \varphi_{N,M} \end{pmatrix}.$$

Each column of this matrix represents a single-particle orbital that is completely specified by its $N$-dimensional vector. For convenience, we will think of the different columns as all properly orthonormalized, which is straightforward to achieve by, for example, modified Gram-Schmidt (see e.g., Zhang 2003, 2013; Motta and Zhang 2018).

The mean-field Hartree-Fock (HF) solution is of course an example of a Slater determinant: $|\phi_{HF}\rangle = \prod_\sigma |\phi_{HF}^\sigma\rangle$, where $|\phi_{HF}^\sigma\rangle$ is defined by a matrix $\Phi_{HF}^\sigma$ whose columns are the $M_\sigma$ lowest HF eigenstates. Similarly, the occupied manifold in a DFT calculation forms a "wave function" which is a Slater determinant.

In standard quantum chemistry (QC) methods, the many-body ground-state wave function is also represented by a sum of Slater determinants. However, there is a key difference between it and the AFQMC representation. In QC methods, the different Slater determinants are orthogonal. As illustrated in the left panel in Fig. 1, each of the determinants is formed by excitations from the HF determinant. In other words, each $|\phi\rangle$ on the right-hand side of Eq. (4) is given by a set of $M$ molecular orbitals (MOs), and the corresponding matrix is formed by orthonormal *unit vectors*. In contrast, in AFQMC the different Slater determinants on the right-hand side of Eq. (4) are not orthogonal: $\langle\phi'|\phi\rangle \neq 0$. They are obtained by rotations of the occupied orbitals using one-body Hamiltonians involving random auxiliary fields (see further details below), as illustrated in the right panel in Fig. 1.

Several properties of non-orthogonal Slater determinants are worth mentioning. The overlap between two of them is given by

$$\langle\phi|\phi'\rangle = \det\left(\Phi^\dagger \Phi'\right). \tag{6}$$

We can define the expectation of an operator $\hat{O}$ with respect to a pair of non-orthogonal Slater determinants:

$$\langle\hat{O}\rangle_{\phi\phi'} \equiv \frac{\langle\phi|\hat{O}|\phi'\rangle}{\langle\phi|\phi'\rangle}, \tag{7}$$

for instance, single-particle Green's function $G_{ij} \equiv \langle c_i c_j^\dagger\rangle_{\phi\phi'}$:

$$G_{ij} \equiv \frac{\langle\phi|c_i c_j^\dagger|\phi'\rangle}{\langle\phi|\phi'\rangle} = \delta_{ij} - [\Phi'(\Phi^\dagger \Phi')^{-1}\Phi^\dagger]_{ij}. \tag{8}$$
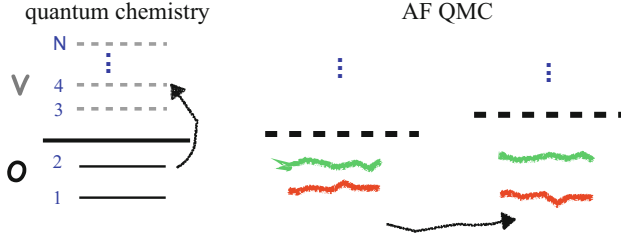
**Fig. 1** Schematic illustration of the connection and difference between quantum chemistry (QC) approaches and AFQMC. A fictitious system, with $M_\uparrow = 2$, $M_\downarrow \leq M_\uparrow$ and $N$ basis functions, is shown. Vertical scale indicates single-particle energy. In QC-based methods (left), Slater determinants are constructed using the molecular orbitals from a (restricted-orbital) HF calculation. ("O" denotes occupied, "V" denotes virtual, and the thick line indicates the Fermi level.) All resulting Slater determinants are orthogonal to each other and to the reference HF state. In AFQMC, each Slater determinant in Eq. (4) is sampled by random walk. Each walker $|\phi\rangle$ has only "occupied" orbitals (denoted by the red and green lines) which are rotated during the random walk, under the influence of stochastic auxiliary fields (illustrated by the *wiggly* lines, which move up and down as the walker evolves from one step to the next). The Slater determinants generated in AFQMC are non-orthogonal to each other

Given the Green's function matrix $G$, the general expectation defined in Eq. (7) can be computed for most operators of interest. For example, we can calculate the expectation of a general two-body operator, $\hat{O} = \sum_{ijkl} O_{ijkl} c_i^\dagger c_j^\dagger c_k c_l$, under the definition of Eq. (7):

$$\langle \hat{O} \rangle_{\phi\phi'} = \sum_{ijkl} O_{ijkl}(G'_{jk}G'_{il} - G'_{ik}G'_{jl}), \tag{9}$$

where the matrix $G'$ is defined as $G' \equiv I - G$.

A key property of Slater determinants we will invoke is the *Thouless Theorem*: any one-particle operator $\hat{B}$ of the form

$$\hat{B} = \exp\left( \sum_{ij} c_i^\dagger U_{ij} c_j \right), \tag{10}$$

when acted on a Slater determinant, simply leads to another Slater determinant (Hamann and Fahy 1990), i.e.,

$$\hat{B}|\phi\rangle = \hat{\phi}_1'^\dagger \hat{\phi}_2'^\dagger \cdots \hat{\phi}_M'^\dagger |0\rangle \equiv |\phi'\rangle \tag{11}$$

with $\hat{\phi}_m'^\dagger = \sum_j c_j^\dagger \Phi'_{jm}$ and $\Phi' \equiv e^U \Phi$, where $U$ is a square matrix whose elements are given by $U_{ij}$ and $B \equiv \exp(U)$ is therefore an $N \times N$ square matrix as well. In other words, the operation of $\hat{B}$ on $|\phi\rangle$ simply involves multiplying an $N \times N$ matrix to the $N \times M$ matrix representing the Slater determinant.

There are several generalizations of the formalism we have discussed which extends the capability and/or accuracy of the AFQMC framework. These can be thought of as generalizing one or both of the Slater determinants in Eqs. (6), (7), and (8). From the viewpoint of AFQMC, as we shall discuss below, the"bra" in these equations represents the trial wave function, and the "ket" represents the random walker:

- The first generalization is to replace $\langle\phi|$ by a projected Bardeen-Cooper-Schrieffer (BCS) wave function, that is, to use a projected BCS as a trial wave function, which can be advantageous for systems with pairing order. The corresponding overlap, Green functions, and two-body mixed expectations have been worked out (Carlson et al. 2011).
- The second is to have both $\langle\phi|$ and $|\phi'\rangle$ in generalized HF (GHF) form, which is necessary to treat systems with spin-orbit coupling (SOC). The required modification to the formalism outlined above is given by Rosenberg et al. (2017).
- The third generalization is to have both sides in Hartree-Fock-Bogoliubov (HFB) form, for example, to treat Hamiltonians with pairing fields. This will also be useful when using AFQMC as an impurity solver in which the embedding induces pairing order. The corresponding AFQMC formalism has been described (Shi and Zhang 2017).

## 2.2    Ground-State Projection

Most ground-state quantum MC (QMC) methods are based on iterative projection:

$$|\Psi_0\rangle \propto \lim_{\tau\to\infty} e^{-\tau\hat{H}}|\Psi_T\rangle; \tag{12}$$

that is, the ground state $|\Psi_0\rangle$ of a many-body Hamiltonian $\hat{H}$ can be projected from any known trial state $|\Psi_T\rangle$ that satisfies $\langle\Psi_T|\Psi_0\rangle \neq 0$. In a numerical method, the limit can be obtained iteratively by

$$|\Psi^{(n+1)}\rangle = e^{-\Delta\tau\hat{H}}|\Psi^{(n)}\rangle, \tag{13}$$

where $|\Psi^{(0)}\rangle = |\Psi_T\rangle$. Ground-state expectation $\langle\hat{O}\rangle$ of a physical observable $\hat{O}$ is given by

$$\langle\hat{O}\rangle = \lim_{n\to\infty} \frac{\langle\Psi^{(n)}|\hat{O}|\Psi^{(n)}\rangle}{\langle\Psi^{(n)}|\Psi^{(n)}\rangle}. \tag{14}$$

For example, the ground-state energy can be obtained by letting $\hat{O} = \hat{H}$. A so-called mixed estimator exists, however, which is exact for the energy (or any other $\hat{O}$ that commutes with $\hat{H}$) and can lead to considerable simplifications in practice:

$$E_0 = \lim_{n \to \infty} \frac{\langle \Psi_T | \hat{H} | \Psi^{(n)} \rangle}{\langle \Psi_T | \Psi^{(n)} \rangle}. \tag{15}$$

QMC methods carry out the iteration in Eq. (13) by Monte Carlo (MC) sampling. The difference between different classes of methods amounts primarily to the space that is used to represent the wave function or density matrix and to carry out the integration. The AFQMC methods work in second quantized representation and in an auxiliary-field space, while Green's function Monte Carlo (GFMC) or diffusion Monte Carlo (DMC) works in first-quantized representation and in electron coordinate space (Kalos et al. 1974; Foulkes et al. 2001). The full-configuration interaction QMC (FCIQMC) (Booth et al. 2009) works in orthogonal Slater determinant space as in QC methods.

Let us assume that $|\Psi_T\rangle$ is of the form of a single Slater determinant or a linear combination of Slater determinants, as in Eq. (4). The operation of $e^{-\tau \hat{H}_1}$ on a Slater determinant simply yields another determinant, per Thouless theorem. The ground-state projection would therefore turn into the propagation of a single Slater determinant if it were somehow possible to write the two-body propagator $e^{-\tau \hat{H}_2}$ as the exponential of a one-body operator.

The above is realized in independent-electron theories. In the HF approximation, $\hat{H}_2$ is replaced by one-body operators times expectations with respect to the current Slater determinant wave function, schematically:

$$c_i^\dagger c_j^\dagger c_k c_l \to c_i^\dagger c_l \langle c_j^\dagger c_k \rangle - c_i^\dagger c_k \langle c_j^\dagger c_l \rangle. \tag{16}$$

(A decomposition that includes pairing is also possible, leading to a Hartree-Fock-Bogoliubov calculation.) In the local density approximation (LDA) in DFT, $\hat{H}_2$ is replaced by $\hat{H}_{\text{LDA}} = \hat{H}_1 + \hat{V}_{xc}$, where $\hat{V}_{xc}$ contains the density operator in real-space, with matrix elements given by the exchange-correlation functional which is computed with the local density from the current Slater determinant in the self-consistent process. In both these cases, an iterative procedure can be used, following Eq. (13), to project out the solution to the approximate Hamiltonians, as an imaginary-time evolution of a single Slater determinant (Zhang and Ceperley 2008). This is illustrated by the blue line in Fig. 2. Note that this procedure is formally very similar to time-dependent HF or time-dependent DFT (TDDFT), except for the distinction of imaginary versus real time.

## 2.3    Hubbard-Stratonovich Transformation

Suppose that $\hat{H}_2$, the two-body part in the Hamiltonian in Eq. (2), can be written as a sum of squares of one-body operators:
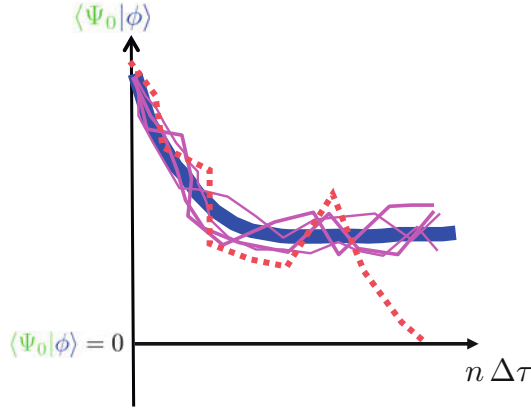
**Fig. 2** Illustration of the iterative imaginary-time projection to the ground state. The overlap of the Slater determinants with a test wave function (e.g., the exact ground-state $|\Psi_0\rangle$) is plotted vs. imaginary time $n\Delta\tau$. The thick blue line indicates a projection using $e^{-\Delta\tau \hat{H}_{\text{LDA}}(\phi^{(n)})}$ which converges to the LDA ground state (or a local minimum). The wiggly magenta lines indicate an AFQMC projection which captures the many-body effect beyond LDA as a stochastic linear superposition. The propagator is obtained by expanding the two-body part of the $\hat{H}$, namely, $\hat{H}_2 - \hat{V}_{xc}$, by a Hubbard-Stratonovich transformation as discussed in the next sections. The dotted redline indicates a path which can lead to a sign problem (Sect. 3.2)

$$\hat{H}_2 = \frac{1}{2} \sum_{\gamma=1}^{N_\gamma} \lambda_\gamma \hat{v}_\gamma^2, \tag{17}$$

where $\lambda_\gamma$ is a constant, $\hat{v}_\gamma$ is a one-body operator similar to $\hat{H}_1$, and $N_\gamma$ is an integer. We can then apply the Hubbard-Stratonovich (HS) transformation to each term

$$e^{-\frac{\Delta\tau}{2} \lambda \hat{v}^2} = \int_{-\infty}^{\infty} dx \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} e^{x\sqrt{-\Delta\tau\lambda}\,\hat{v}}, \tag{18}$$

where $x$ is an auxiliary-field variable. The constant in front of $\hat{v}$ in the exponent on the right-hand side can be real or imaginary, depending on the sign of $\lambda$. The key is that the quadratic form (in $\hat{v}$) on the left is replaced by a linear one on the right. There are various ways to achieve the decomposition in Eq. (17) for a general two-body term (Negele and Orland 1998). Below we outline the two most commonly applied cases in electronic structure: (**a**) *with plane-wave basis* and (**b**) for a more dense matrix $V_{ijkl}$ resulting from a *general basis set* such as Gaussians in QC.

In a *plane-wave basis*, the two-body part is the Fourier transform of $1/|\mathbf{r}_m - \mathbf{r}_n|$ (Suewattana et al. 2007):

$$\hat{H}_2 \rightarrow \frac{1}{2\Omega} \sum_{i,j,k,l} \frac{4\pi}{|\mathbf{G_i} - \mathbf{G_k}|^2} c_i^\dagger c_j^\dagger c_l c_k \delta_{\mathbf{G_i}-\mathbf{G_k},\mathbf{G_l}-\mathbf{G_j}} \delta_{\sigma_i,\sigma_k} \delta_{\sigma_j,\sigma_l}, \tag{19}$$

where $\{\mathbf{G_i}\}$ are plane-wave wave-vectors, $\Omega$ is the volume of the supercell, and $\sigma$ denotes spin. Let us use $\mathbf{Q} \equiv \mathbf{G_i} - \mathbf{G_k}$ and define a density operator in momentum space:

$$\hat{\rho}(\mathbf{Q}) \equiv \sum_{\mathbf{G},\sigma} c^{\dagger}_{\mathbf{G}+\mathbf{Q},\sigma} c_{\mathbf{G},\sigma}, \tag{20}$$

where the sum is over all $\mathbf{G}$ vectors which allow both $\mathbf{G}$ and $\mathbf{G}+\mathbf{Q}$ to fall within the predefined kinetic energy cutoff in the plane-wave basis. The two-body term in Eq. (19) can then be manipulated into the form

$$\hat{H}_2 \to \sum_{\mathbf{Q}\neq\mathbf{0}} \frac{\pi}{\Omega Q^2} \left[ \hat{\rho}^{\dagger}(\mathbf{Q})\,\hat{\rho}(\mathbf{Q}) + \hat{\rho}(\mathbf{Q})\,\hat{\rho}^{\dagger}(\mathbf{Q}) \right], \tag{21}$$

where the sum is over all $\mathbf{Q}$'s except $\mathbf{Q} = 0$, since in Eq. (19) the $\mathbf{G_i} = \mathbf{G_k}$ term is excluded due to charge neutrality, and we have invoked $\rho^{\dagger}(\mathbf{Q}) = \rho(-\mathbf{Q})$. By making linear combinations of $\left[(\rho^{\dagger}(\mathbf{Q}) + \rho(\mathbf{Q})\right]$ and $\left[(\rho^{\dagger}(\mathbf{Q}) - \rho(\mathbf{Q})\right]$ terms, we can then readily write the right-hand side in Eq. (21) in the desired square form of Eq. (17) (Suewattana et al. 2007).

With a *general basis* such as Gaussians yielding matrix elements given in Eq. (3), the most straightforward way to decompose $\hat{H}_2$ is through a direct diagonalization (Al-Saidi et al. 2006, 2007; Zhang 2013). However, this is computationally costly. A modified Cholesky decomposition leads to $\mathcal{O}(N)$ fields (Purwanto et al. 2011; Motta and Zhang 2018). This approach, which has been commonly used in AFQMC for molecular systems with Gaussian basis sets and for downfolded Hamiltonians (Ma et al. 2015), proceeds as follows. Let us cast $V_{ijkl}$ in the form of a two-index matrix by introducing the compound indices $\mu = (i, l)$ and $\nu = (j, k)$: $V_{\mu\nu} = V_{(i,l),(j,k)} = V_{ijkl}$. The symmetric positive semidefinite matrix $V_{\mu\nu}$ is decomposed using a recursive modified Cholesky algorithm (Koch et al. 2003; Aquilante et al. 2010), to yield

$$V_{\mu\nu} = \sum_{\gamma=1}^{N_\gamma} L_\mu^\gamma L_\nu^\gamma + \Delta_{\mu\nu}^{(N_\gamma)}, \tag{22}$$

where $\Delta_{\mu\nu}^{(N_\gamma)}$ is the residual error at the $N_\gamma$-th iteration. The iterative procedure is repeated until all matrix elements of the residual matrix are less than some predefined tolerance $\delta$:

$$\left| V_{\mu\nu} - V_{\mu\nu}^{(N_{\mathrm{CD}})} \right| = \left| \Delta_{\mu\nu}^{(N_{\mathrm{CD}})} \right| \leq \delta. \tag{23}$$

For molecular calculations, typical values of $\delta$ range between $10^{-4}$ and $10^{-6}$ in atomic units (Motta and Zhang 2018). Using the $N_{\mathrm{CD}}$ Cholesky vectors, we can rewrite the two-body part of the Hamiltonian

$$\hat{H}_2 \rightarrow \frac{1}{2} \sum_{\gamma=1}^{N_{\text{CD}}} \left( \sum_{il} L_{\mu(i,l)}^{\gamma} c_i^{\dagger} c_l \right) \left( \sum_{jk} L_{\nu(j,k)}^{\gamma} c_j^{\dagger} c_k \right) + \mathcal{O}(\delta). \tag{24}$$

Hence the form in Eq. (17) is realized, with $\hat{v}_{\gamma} = \sum_{il} L_{\mu(i,l)}^{\gamma} c_i^{\dagger} c_l$.

Different forms of the HS transformation can affect the performance of the AFQMC method. For example, it is useful to subtract a mean-field "background" from the two-body term prior to the decomposition (Baer et al. 1998; Purwanto and Zhang 2005; Al-Saidi et al. 2006). The idea is that using the HS to decompose any constant shifts in the two-body interaction will necessarily result in more statistical noise. In fact, it has been shown (Shi and Zhang 2013; Motta and Zhang 2018) that the mean-field background subtraction can not only impact the statistical accuracy but also lead to different quality of approximations under the constrained path methods that we discuss in the next section.

If we denote the collection of auxiliary fields by **x** and combine one-body terms from $\hat{H}_1$ and $\hat{H}_2$, we obtain the following compact representation of the outcome of the HS transformation:

$$e^{-\Delta\tau\hat{H}} = \int d\mathbf{x}\, p(\mathbf{x}) \hat{B}(\mathbf{x}), \tag{25}$$

where $p(\mathbf{x})$ is a probability density function (PDF), for example, a multidimensional Gaussian. The propagator $\hat{B}(\mathbf{x})$ in Eq. (25) has a *special form*, namely, it is a product of operators of the type in Eq. (10), with $U_{ij}$ depending on the auxiliary field. The matrix representation of $\hat{B}(\mathbf{x})$ will be denoted by $B(\mathbf{x})$.

Note that the matrix elements of $B(\mathbf{x})$ can become complex, for example, when $\lambda$ in Eq. (18) is positive, which occurs in both of the forms discussed above. Sometimes we will refer to this situation as having complex auxiliary fields, but it should be understood that terms in the PDF of the HS transformation and $\hat{B}(\mathbf{x})$ can be rearranged, and the relevant point is whether the Slater determinant has matrix elements which are real or complex, as further discussed in the next section.

In essence, the HS transformation replaces the two-body interaction by one-body interactions with a set of random external auxiliary fields. In other words, it converts an interacting system into many *noninteracting* systems living in fluctuating external auxiliary fields. The sum over all configurations of auxiliary fields recovers the interaction.

## 3  Ground-State AFQMC Methods

### 3.1  Free-Projection AFQMC

We first briefly describe the ground-state AFQMC method without any constraints. Our goal is to illustrate the essential ideas, in a way which will facilitate our ensuing discussions and help introduce the constrained path approximation and the

framework for the general AFQMC methods that control the sign/phase problem. We will not go into details, which are described in the literature.

We write the usual path-integral and Metropolis form explicitly here to show its connection to the open-ended random walk approach. Ground-state expectation $\langle \hat{O} \rangle$ can be computed with Eqs. (13) and (25). The denominator is

$$
\langle \psi^{(0)} | e^{-n\Delta\tau\hat{H}} e^{-n\Delta\tau\hat{H}} | \psi^{(0)} \rangle
$$

$$
= \int \langle \psi^{(0)} | \left[ \prod_{l=1}^{2n} d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)}) \hat{B}(\mathbf{x}^{(l)}) \right] | \psi^{(0)} \rangle
$$

$$
= \int \left[ \prod_l d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)}) \right] \det\left( [\Psi^{(0)}]^\dagger \prod_l B(\mathbf{x}^{(l)}) \Psi^{(0)} \right). \tag{26}
$$

In the standard ground-state AFQMC method (Sugiyama and Koonin 1986; Sorella et al. 1989; Blankenbecler et al. 1981), a value of $n$ is first chosen and is kept fixed throughout the calculation. If we use $X$ to denote the collection of the auxiliary fields $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(2n)}\}$ and $D(X)$ to represent the integrand in Eq. (26), we can write the expectation value of Eq. (14) as

$$
\langle \hat{O} \rangle = \frac{\int \langle \hat{O} \rangle_{LR} D(X) \, dX}{\int D(X) \, dX} = \frac{\int \langle \hat{O} \rangle_{LR} |D(X)| \Theta(X) \, dX}{\int |D(X)| \Theta(X) \, dX}, \tag{27}
$$

where

$$
\Theta(X) \equiv D(X)/|D(X)| \tag{28}
$$

measures the phase of $D(X)$, which reduces to a sign when the overlap, $D(X)$, is real along all paths $\{X\}$. The expectation for a given $X$, as defined in Eq. (7), is:

$$
\langle \hat{O} \rangle_{LR} \equiv \frac{\langle \phi_L | \hat{O} | \phi_R \rangle}{\langle \phi_L | \phi_R \rangle} \tag{29}
$$

with

$$
\langle \phi_L | = \langle \psi^{(0)} | \hat{B}(\mathbf{x}^{(2n)}) \hat{B}(\mathbf{x}^{(2n-1)}) \cdots \hat{B}(\mathbf{x}^{(n+1)})
$$

$$
| \phi_R \rangle = \hat{B}(\mathbf{x}^{(n)}) \hat{B}(\mathbf{x}^{(n-1)}) \cdots \hat{B}(\mathbf{x}^{(1)}) | \psi^{(0)} \rangle,
$$

which are both Slater determinants.

$D(X)$ and $\langle \phi_L |$ and $| \phi_R \rangle$ are completely specified by the path $X$ in auxiliary-field space, given $|\Psi^{(0)}\rangle$. The expectation in Eq. (27) is thus a many-dimensional

integration which can be evaluated by standard MC techniques. Often the Metropolis algorithm (Kalos and Whitlock 1986) is used to sample auxiliary fields $X$ from $|D(X)|$. We will refer to this as free-projection (in contrast with a constrained path calculations). There are special Hamiltonians (e.g., repulsive Hubbard model at half-filling) where special symmetry makes the sign problem absent. In those situations, the Metropolis approach described above is very effective and is the standard approach. It should be mentioned that in those cases, an infinite variance problem arises which must be controlled (Shi and Zhang 2016).

We carry out free-projection calculations with an open-ended random walk (Zhang et al. 1997; Baer et al. 1998) instead of using Metropolis sampling outlined above. For free-projection calculations, the open-ended approach has no real advantage. However, when a sign or phase problem is present, it is difficult to implement a constraint to control the problem in the Metropolis framework, because of ergodicity issues (Fahy and Hamann 1990; Zhang et al. 1997). The open-ended random walk framework avoids the difficulty and is straightforward to project to longer imaginary time in order to approach the ground state. Moreover, when we carry out constraint release (Shi and Zhang 2013), the formalism will rely on the open-ended random walk. These points will become clear after we illustrate the phase problem in electronic structure calculations below and discuss how the constraint can be formulated. The structure of the open-ended random walk is illustrated in Fig. 3.
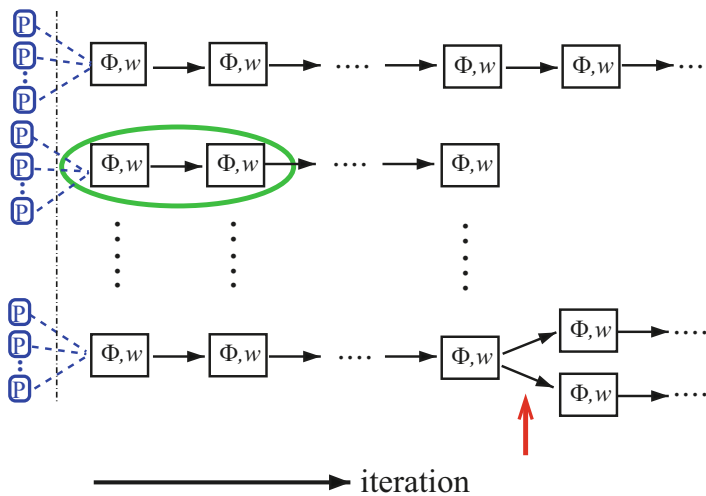


**Fig. 3** Schematic illustration of the AFQMC method. Each box is a random walker, with $|\phi\rangle$ the stochastic Slater determinant and $w$ its weight. A step (green oval) is similar to one SCF (self-consistent-field) step in LDA. The red arrow indicates a population control, where birth/death can occur. This structure allows for exceptional capacity for scaling on parallel computers. Multiple walkers can reside on one processor (blue "P" box), or each walk can be split over processors for large problems

## 3.2    Constrained Path AFQMC

As mentioned, a sign/phase problem occurs in the free-projection AFQMC, except for special cases where the single-particle propagator $\hat{B}(\mathbf{x})$ satisfies particular symmetries (see, e.g., Wei et al. 2016). In these cases, $\Theta(X)$ (defined in Eq. (28)) vanishes, and $D(x)$ is real and nonnegative. Absent such special circumstances, a sign problem arises if $\hat{B}(\mathbf{x})$ is real, and a phase problem arises if $\hat{B}(\mathbf{x})$ is complex. As mentioned, the Coulomb interaction in $V_{\text{int}}$ leads to a phase problem in molecules and solids. In this section we discuss the constrained-path (CP) AFQMC, which for electronic systems has often been referred to as the phaseless or phase-free approximation (Zhang and Krakauer 2003; Purwanto and Zhang 2004).

For real $\hat{B}(\mathbf{x})$ (e.g., Hubbard-type of short-range repulsive interactions decoupled with spin form of HS transformation), the sign problem occurs because of the fundamental symmetry between the fermion ground-state $|\Psi_0\rangle$ and its negative $-|\Psi_0\rangle$ (Zhang 1999a; Zhang and Kalos 1991). For any ensemble of Slater determinants $\{|\phi\rangle\}$ which gives a MC representation of the ground-state wave function, as in Eq. (4), this symmetry implies that there exists another ensemble $\{-|\phi\rangle\}$ which is also a correct representation. In other words, the Slater determinant space can be divided into two degenerate halves ($+$ and $-$) whose bounding surface $\mathscr{N}$ is defined by $\langle\Psi_0|\phi\rangle = 0$. This dividing surface is unknown. (In the cases with special symmetry mentioned above, the two sides separated by the surface are both positive. This has to do with the over-complete nature of the non-orthogonal Slater determinant space in AFQMC. A particular form of $\hat{B}(\mathbf{x})$ can pick out only a part of the space which can be nonnegative.)

The idea of the $\pm$-symmetry can be seen from Fig. 2, where the dotted red line indicates a walker reaching the surface $\mathscr{N}$, which will have a finite probability of occurring in a random walk, unless completely excluded by the dynamics. Once it does, it can, in general, freely sample the two families of solutions which are symmetric about the horizontal axis (above and below). A more detailed illustration and discussion of the sign problem can be found in Zhang (1999b, 2013). The idea of the phase problem is illustrated in Fig. 4. The complex plane now replaces the vertical axis in Fig. 2, denoting the overlap of a random walker $|\phi\rangle$ with the (hypothetically) known exact wave function or the trial wave function $\langle\Psi_T|$.

Cancelation schemes can partially alleviate the problem, as demonstrated in coordinate space (Zhang and Kalos 1991; Diedrich and Anderson 1992) and in Fock space (Booth et al. 2009). To fully stabilize the calculation and restore polynomial scaling, however, an approximation has been necessary. To date, the most effective and accurate method to achieve this has been the constrained path approach.

The method begins with a generalized similarity transformation in the spirit of the importance-sampling transformation. To make the derivation more concrete, we will use an explicit form of the HS transformation and write the two-body propagator that results from Eq. (21) or (24) as $\int e^{-\frac{\mathbf{x}^2}{2}} e^{\mathbf{x}\cdot\hat{\mathbf{v}}} d\mathbf{x}$, where $\mathbf{x}$ is an $N_\gamma$-dimensional vector as given by Eq. (25) (with $\gamma$ labeling the auxiliary fields) and
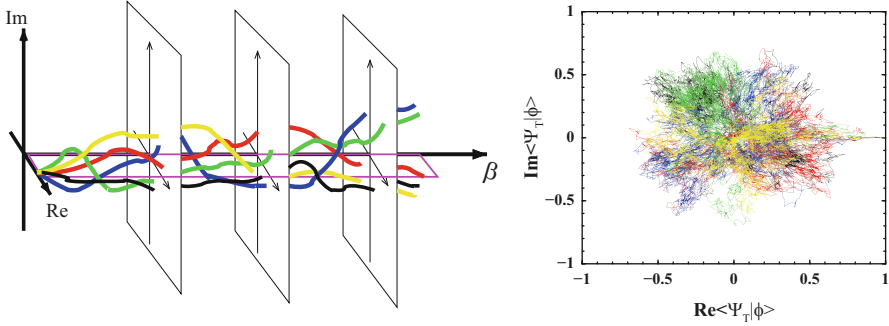
**Fig. 4** Schematic illustration of the phase problem and constraints to control it. The **left panel** shows, as a function of projection time $\beta \equiv n\Delta\tau$, trajectories of five walkers (shown as five different colors) characterized by the real (Re) and imaginary (Im) parts of their overlap with the ground-state wave function. The **right panel** shows the walker distribution integrated over imaginary time, i.e., the different frames in the left panel stacked together along $\beta$. The phase problem occurs because the initial phase "coherence" of the random walkers rapidly deteriorates with $\beta$, as they become uniformly distributed in the Re-Im-plane. The idea of the phase constraint (Zhang and Krakauer 2003) is to apply a gauge transformation such that confining the random walk in the single magenta plane (left) is a good approximation

$\hat{\mathbf{v}} = \{\hat{v}_\gamma\}$ denotes the collection of one-body operators. We introduce a shift to obtain an alternative propagator:

$$\int e^{-\frac{\mathbf{x}^2}{2}} e^{\mathbf{x}\cdot\bar{\mathbf{x}} - \frac{\bar{\mathbf{x}}^2}{2}} e^{(\mathbf{x}-\bar{\mathbf{x}})\cdot\hat{\mathbf{v}}} d\mathbf{x} \,, \tag{30}$$

which is exact for any choice of the shift $\bar{\mathbf{x}}$, including complex shifts.

We recall that the random walk is supposed to lead to a MC sampling of the coefficient $\alpha_\phi$ in Eq. (4):

$$|\Psi_0\rangle \doteq \sum_{\{\phi\}} w_\phi |\phi\rangle \,. \tag{31}$$

The sum in Eq. (31), which is over the population of walkers after an "equilibration" portion of the open-ended random walk has been discarded, is in a Monte Carlo sense, and is typically much smaller than the sum in Eq. (4). The weight of each walker $|\phi\rangle$, $w_\phi$, can be thought of as 1 (all walkers with equal weight); it is allowed to fluctuate only for practical (efficiency) consideration.

Using the idea of importance sampling, we seek to replace Eq. (31) by the following to sample Eq. (4):

$$|\Psi_0\rangle = \sum_\phi w_\phi \frac{|\phi\rangle}{\langle\Psi_T|\phi\rangle} \,, \tag{32}$$

where any overall phase of the walker $|\phi\rangle$ is canceled in the numerator and denominator on the right-hand side (Zhang and Krakauer 2003). This implies a modification to the propagator in Eq. (30):

$$\int \langle \Psi_T | \phi'(\mathbf{x}) \rangle e^{-\frac{\mathbf{x}^2}{2}} e^{\mathbf{x}\bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2} e^{(\mathbf{x}-\bar{\mathbf{x}})\cdot \hat{\mathbf{v}}} \frac{1}{\langle \Psi_T | \phi \rangle} d\mathbf{x}, \tag{33}$$

where $|\phi'(\mathbf{x})\rangle = e^{(\mathbf{x}-\bar{\mathbf{x}})\cdot \hat{\mathbf{v}}}|\phi\rangle$ and the trial wave function $|\Psi_T\rangle$ represents the best guess to $|\Psi_0\rangle$. Let us define the following shorthand:

$$\bar{\mathbf{v}} \equiv -\frac{\langle \Psi_T | \hat{\mathbf{v}} | \phi \rangle}{\langle \Psi_T | \phi \rangle} \sim \mathcal{O}(\sqrt{\Delta\tau}); \quad \bar{\mathbf{v}^2} \equiv \frac{\langle \Psi_T | \hat{\mathbf{v}}^2 | \phi \rangle}{\langle \Psi_T | \phi \rangle} \sim \mathcal{O}(\Delta\tau). \tag{34}$$

We can then evaluate the ratio $\langle \Psi_T | \phi'(\mathbf{x}) \rangle / \langle \Psi_T | \phi \rangle$ in Eq. (33) by expanding the propagator (Moskowitz et al. 1982; Zhang and Krakauer 2003; Purwanto and Zhang 2004) to $\mathcal{O}(\tau)$, to obtain:

$$\frac{\langle \Psi_T | \phi'(\mathbf{x}) \rangle}{\langle \Psi_T | \phi \rangle} e^{\mathbf{x}\cdot\bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2} \doteq \exp\left[ -(\mathbf{x}-\bar{\mathbf{x}})\cdot\bar{\mathbf{v}} + \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^2\bar{\mathbf{v}^2} - \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^2\bar{\mathbf{v}}^2 + \mathbf{x}\cdot\bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2 \right]. \tag{35}$$

The optimal choice of the shift $\bar{\mathbf{x}}$, which we shall refer to as a force bias, minimizes the fluctuation of Eq. (35) with respect to $\mathbf{x}$, and it is straightforward to show that it is $\bar{\mathbf{x}} = \bar{\mathbf{v}}$. With this choice, Eq. (33) can be written approximately as (Zhang 2013)

$$\int e^{-\frac{\mathbf{x}^2}{2}} e^{(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}} e^{\frac{\bar{\mathbf{v}}^2}{2}} d\mathbf{x}. \tag{36}$$

Restoring $\hat{H}_1$, we obtain the complete propagator:

$$\int e^{-\frac{\mathbf{x}^2}{2}} \exp\left[ -\frac{\Delta\tau \hat{H}_1}{2} \right] \exp\left[ (\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}} \right] \exp\left[ -\frac{\Delta\tau \hat{H}_1}{2} \right] \exp[-\Delta\tau E_L(\phi)] \, d\mathbf{x}, \tag{37}$$

where $E_L$ is the local energy, the mixed estimate of the Hamiltonian:

$$E_L(\phi) \equiv \frac{\langle \Psi_T | \hat{H} | \phi \rangle}{\langle \Psi_T | \phi \rangle}. \tag{38}$$

In the limit of an exact $|\Psi_T\rangle$, $E_L$ is a *real* constant, and the weight of each walker remains real. The mixed estimate for the energy from Eq. (15) is phaseless:

$$E_0^c = \frac{\sum_\phi w_\phi E_L(\phi)}{\sum_\phi w_\phi}. \tag{39}$$

With a general $|\Psi_T\rangle$ which is not exact, a natural approximation is to replace $E_L$ in Eq. (37) by its real part, Re$E_L$. The same replacement is then necessary in Eq. (39).

When $\hat{B}(\mathbf{x})$ (i.e., $\hat{\mathbf{v}}$) is real, this formalism reduces to the so-called constrained-path approximation (Zhang et al. 1997). Regardless of whether $\hat{\mathbf{v}}$ is real, the shift $\bar{\mathbf{x}}$ diverges as the random walk in the complex plane (see the right panel of Fig. 4) approaches the origin, i.e., as $\langle\Psi_T|\phi'\rangle \to 0$. The effect of the divergence is to move the walker away from the origin. With a *real* $\hat{\mathbf{v}}$, the random walkers move only on the real axis. If they are initialized to have positive overlaps with $|\Psi_T\rangle$, $\bar{\mathbf{x}}$ will ensure that the overlaps remain positive throughout the random walk.

For a general case with a complex $\hat{\mathbf{v}}$, however, the formalism above by itself is not sufficient to remove the phase problem. To see this we consider the phase of $\langle\Psi_T|\phi'(\mathbf{x}-\bar{\mathbf{x}})\rangle/\langle\Psi_T|\phi\rangle$, which we denote by $\Delta\theta$. In general, $\Delta\theta \sim \mathcal{O}(-\mathbf{x}\text{Im}(\bar{\mathbf{x}}))$ is non-zero. This means that the walkers will undergo a random walk in the complex plane. At large $\beta$ they will therefore populate the complex plane symmetrically, independent of their initial positions. This is illustrated in the right panel of Fig. 4, which shows $\langle\Psi_T|\phi\rangle$ for a three-dimensional jellium model with two electrons at $r_s = 10$ for a total projection time of $\beta = 250$ (taken from Zhang 2013). The random walk is "rotationally invariant" in the complex plane, resulting in a vanishing signal-to-noise ratio asymptotically, even though the walkers are all real initially with $\langle\Psi_T|\phi^{(0)}\rangle = 1$. An alternative but related way to state the problem is that, despite the divergence of $\bar{\mathbf{x}}$, the buildup of a finite density at the origin of the complex plane cannot be prevented, unlike in the one-dimensional situation (real $\langle\Psi_0|\phi\rangle$, sign problem). Near the origin the local energy $E_L$ diverges, which causes diverging fluctuations in the weights of walkers when the density does not vanish.

Thus the second ingredient of the constraint for the phase problem is to project the random walk back to "one-dimension." This is done by reducing the weight of the walker in each step by the angular deviation of the overlap in the complex plane:

$$w_{\phi'} \to w_{\phi'} \max\{0, \cos(\Delta\theta)\}. \qquad (40)$$

A prerequisite for this approximation to work well is the importance-sampling transformation, which has eliminated the leading order in the overall phase of $|\phi\rangle$ in the propagator in Eq. (35). Given the transformation, several alternative forms to the projection in Eq. (40) were found to give similar accuracy (Zhang and Krakauer 2003; Zhang et al. 2005; Purwanto and Zhang 2005; Zhang 2013).

We can now summarize each step in the constrained path AFQMC formalism as follows. For each random walker $|\phi\rangle$ in the current population $\{|\phi\rangle, w_\phi\}$,

**(a)** Sample $\mathbf{x}$ and propagate the walker to $|\phi'\rangle$

$$|\phi\rangle \to |\phi'\rangle = \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right] \exp\left[(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}\right] \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right]|\phi\rangle, \qquad (41)$$

**(b)** Update the weight of the walker

$$w_\phi \to w_{\phi'} = w_\phi \exp\left[-\Delta\tau \cdot \text{Re}\left(E_L(\phi') + E_L(\phi)\right)/2\right] \cdot \max\{0, \cos(\Delta\theta)\}.$$
$$(42)$$

Walkers so generated represent the ground-state wave function with importance sampling, in the sense of Eq. (32).

For additional technical details, we refer the reader to Zhang (2013) and Motta and Zhang (2018), and references therein, for example, re-orthogonalization procedures (White et al. 1989; Zhang 2003) to stabilize the walkers against numerical roundoff errors during the propagation, population control (Umrigar et al. 1993; Zhang et al. 1997) to regularize the branching process, a hybrid alternative (Purwanto et al. 2009a) to the local energy formalism in Eq. (36) to reduce computational cost in evaluating $E_l$, correlated sampling (Shee et al. 2017), constraint release (Shi and Zhang 2013), etc.

## 3.3   Back-Propagation for Observables and Correlation Functions

To calculate a correlation function or the expectation value of an observable which does not commute with the Hamiltonian, the mixed estimate in Eq.(15) is biased, and the full estimator in Eq. (14) needs to be computed. In the path-integral form in Eq. (27), this is straightforward. With the open-ended random walks, however, it is slightly more involved. A back-propagation (BP) technique (Zhang et al. 1997; Purwanto and Zhang 2004; Motta and Zhang 2017) is employed.

The idea of the BP is to create two coupled populations to represent the bra and ket in Eq. (14), respectively. Because the population in the random walk is importance-sampled, two independent populations which are uncoupled would lead to large fluctuations in the estimator after the importance functions have been "undone" (Purwanto and Zhang 2004). In BP, we choose an iteration $n$ and store the entire population $\{|\phi_k^{(n)}\rangle, w_k^{(n)}\}$, where $k$ labels the walker in the population. As the random walk proceeds from $n$, we keep track of the following two items for each new walker: (1) the sampled auxiliary-field values that led to the new walker from its parent walker and (2) an integer label that identifies the parent. After an additional $m$ iterations, we carry out the back-propagation: For each walker $l$ in the $(n + m)$-th (current) population, we initiate a determinant $\langle\psi_T|$ and act on it with the corresponding propagators but taken in reverse order. The $m$ successive propagators are constructed from the items stored between steps $n + m$ and $n$, with $\exp(-\Delta\tau\hat{H}_1/2)$ inserted where necessary. The resulting determinants $\langle\bar{\phi}_l^{(m)}|$ are combined with its parent from iteration $n$, $|\phi_k^{(n)}\rangle$, to compute $\langle\mathcal{O}\rangle_{\text{BP}}$, where $k$ is the index of the walker at step $n$ from which walker $l$ at step $(n + m)$ descended.

In molecular systems, an improvement over the standard procedure has been proposed (Motta and Zhang 2017). The approach, called back-propagation with path restoration (BP-PRes), allows one to "undo" some of the effect of the constraint in

the forward direction (the cos projection and the omission of the phase in $E_l$ in the weight). This reduced the effect of the constraint, which is applied in the forward direction and does not preserve reversal symmetry in imaginary time. With these advances, accurate observables and atomic forces have been obtained in molecules, paving the way for geometry optimization and ab initio molecular dynamics with AFQMC.

Another recent development in methodology is the computation of imaginary-time correlation functions and excitations. The techniques (Vitali et al. 2016), which have been applied in model systems and ultracold atoms so far, are directly applicable to real materials.

## 4    Illustrative Results

The AFQMC method has been applied to lattice models, realistic solids (using plane-wave basis and pseudo potentials), molecular systems (using Gaussian basis sets), and downfolded model Hamiltonians of real materials (using DFT orbitals as basis sets). The method is just coming into form, and rapid advances in algorithmic development and in applications are ongoing. We briefly mention a few examples here to provide a flavor of how it has been applied to date to tackle problems of electron correlations in materials.

For lattice models, most of the applications involve "only" a sign problem, because of the short-range nature of the interaction. Here the constraint has no $\theta$ projection and reduces to a sign constraint. A large body of results exist, including recent benchmark results (LeBlanc et al. 2015). Systems of $\mathcal{O}(1000)$ electrons have been treated quite routinely. The AFQMC method has demonstrated excellent capabilities and accuracy, illustrating its potential as a general many-body computational paradigm. A key recent development (Qin et al. 2016) is to use the density or density matrix computed from AFQMC as a feedback into a mean-field calculation. The trial wave function $|\Psi_T\rangle$ obtained from the mean-field is then fed back into the AFQMC as a constraint, and a self-consistent constraining condition is achieved. This has led to further improvement in the accuracy and robustness of the calculation (Qin et al. 2016; Zheng et al. 2017).

For molecular systems, a recent review article is available (Motta and Zhang 2018) which describes in more detail the application of AFQMC in quantum chemistry. The formulation of AFQMC with Gaussian basis sets has been extremely valuable. Direct comparisons can be made with high-level QC results, which have provided valuable benchmark information and have been crucial in gauging the AFQMC method as a general approach. Figure 5 illustrates the results on molecules using both plane-wave plus pseudopotentials and Gaussian basis sets. In these calculations we have operated largely in an automated mode, inputting only the DFT or HF solutions as $|\Psi_T\rangle$. This illustrates a potential mode of operation for AFQMC as a "post-processing" approach for molecules and solids where additional accuracy is desired beyond standard DFT.
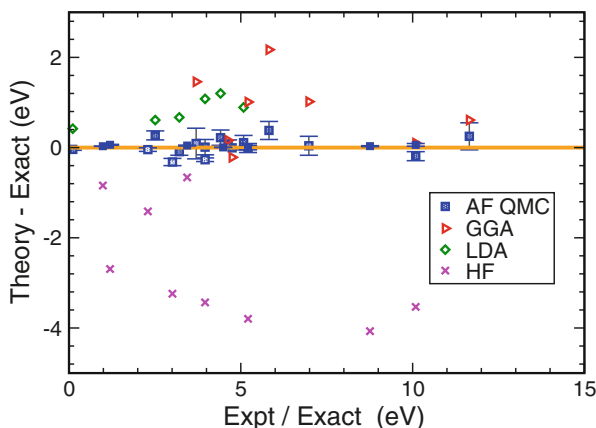
**Fig. 5** Calculated binding energies of molecules compared with experimental values. (Taken from Zhang (2013); Esler et al. 2008). The discrepancy between theory and experiment is plotted. Included are $sp$-bonded molecules, first- and second-row post-$d$ elements, and transition metal oxides. Several different forms of the AFQMC calculations were tested, including all-electron Gaussian basis sets, Gaussian basis with effective-core potentials, and plane wave with pseudopotential. The AFQMC is fed a trial wave function to start, which is taken directly from DFT [with either LDA or the generalized-gradient approximation (GGA) functionals] or HF. The corresponding DFT or HF results are also shown. As can be readily observed, the AFQMC results are in excellent agreement with experiment and significantly improve upon the values from DFT and HF

A benchmark study (Motta et al. 2017) was recently carried out involving a large set of modern many-body methods. AFQMC was among the methods included; consistent with previous findings, the accuracy of AFQMC is found to be comparable to CCSD(T), the gold standard in chemistry (Bartlett and Musiał 2007; Crawford and Schaefer 2000), near equilibrium geometry. For bond breaking, AFQMC was able to maintain systematic accuracy. Large basis sets and system sizes were reached and an accurate equation of state was obtained.

The AFQMC method can be used to study excited states. Excited states distinguished by different symmetry from the ground state can be computed in a manner similar to the ground state. For other excited states, prevention of collapse into the ground state and control of the fermion sign/phase problem are accomplished by a constraint using an excited state trial wave function (Purwanto et al. 2009b). An additional orthogonalization constraint is formulated to use virtual orbitals in solids for band structure calculations (Ma et al. 2013). These constraints are not as "clean" or rigorous as that for the ground state. Use of improved trial wave functions (e.g., multi-determinant $|\Psi_T\rangle$ in molecules) and the imposition of symmetry properties (Shi and Zhang 2013) often lead to improved results. Tests in the challenging case of the $C_2$ molecule yielded spectroscopic constants in excellent agreement with experiment (Purwanto et al. 2009b). In Fig. 6 results from an application in solids are shown for the diamond band structure and for the fundamental band gap in wurtzite ZnO (Ma et al. 2013).
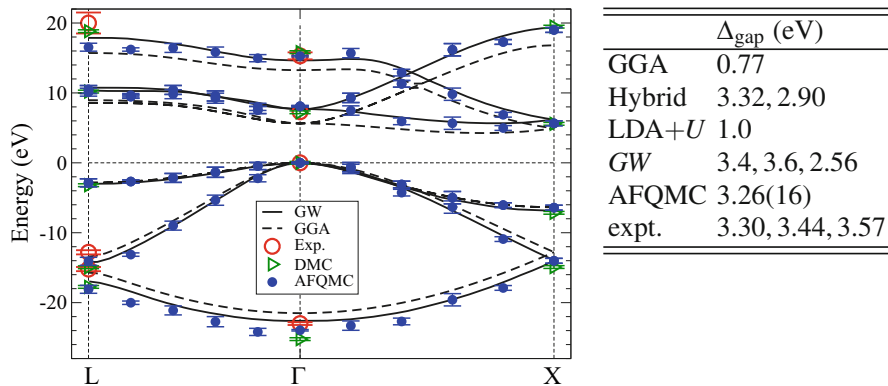
**Fig. 6** Computation of excitations and many-body quasiparticle band structures. (Taken from Ma et al. 2013). The figure presents results on the band gap in diamond. Blue is AFQMC results; $GW$ and DFT band structures are plotted by solid and dashed lines, respectively. Diffusion Monte Carlo (DMC) results at high symmetry points $\Gamma$, $X$, and $L$ are indicated by green triangles. Experimental values are shown as red circles. The table shows the calculated fundamental band gap of wurtzite ZnO, compared with experiment (and three DFT-based methods and $GW$)

Plane-wave calculations in AFQMC can be built on standard plane-wave technologies in DFT calculations, as we have outlined. Norm-conserving pseudopotentials, including multiple-projector pseudopotentials, can be implemented straightforwardly (Ma et al. 2017). In order to reduce the cost of full plane-wave AFQMC calculations, a downfolding approach (Ma et al. 2015) has been developed. The idea is to use Kohn-Sham orbitals (occupied and virtual) as basis sets. The approach is illustrated in Fig. 7. The size of the basis set in the largest calculation on the right, after downfolding, is more than an order of magnitude smaller than the total number of plane waves, leading to large savings in the AFQMC computation.

Figure 8 illustrates an application of AFQMC to the adsorption of Co atoms on graphene. The goal of the study was to determine the stability and magnetic state of the Co adatom as a function of its distance from the graphene sheet. The sensitivity and complexity of the energetics requires a correlated treatment. In addition to serving as a useful benchmark, the computed results provided an explanation for experimental results with Co on free-standing graphene (Virgus et al. 2014). The AFQMC calculation was performed by embedding it in a DFT calculation to extend length scales, as illustrated on the right. After the DFT is performed, a many-body Hamiltonian is generated, using a procedure similar to that of producing a frozen-core Hamiltonian (Purwanto et al. 2013), except the "core" here is actually the "outer" region indicated on the right. (Orbital localization procedures are applied as needed.) The resulting Hamiltonian, which describes the region inside the shaded circle embedded in the environment of the outer region whose orbitals are frozen, is then treated by AFQMC. From a QC perspective, this approach can also be viewed as casting AFQMC as a general "solver' for a (very large) active space (Motta and Zhang 2018).
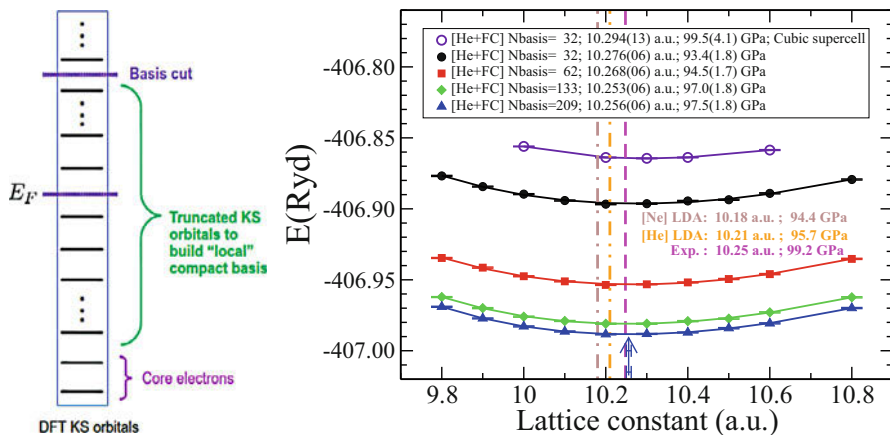
**Fig. 7** Pseudopotential-free calculations in solids, and a simple "down-folding" approach to generate realistic model Hamiltonians. The **left panel** illustrates a scheme to use Kohn-Sham (KS) orbitals obtained from a DFT calculation as basis set for AFQMC. The DFT is performed with a plane-wave basis using a helium-core. Keeping all the KS orbitals, including virtual orbitals, below a certain cutoff ("Basis cut"), we compute the matrix elements with these orbitals as basis, to obtain a many-body Hamiltonian, which is then fed into the AFQMC. In the AFQMC, the KS orbitals corresponding to the neon-core are frozen (Purwanto et al. 2013). In the **right panel**, the calculated equation of state (Ma et al. 2015) is shown for a sequence of "Basis cut" values. The calculated equilibrium lattice constant and bulk modulus are in excellent agreement with experiment
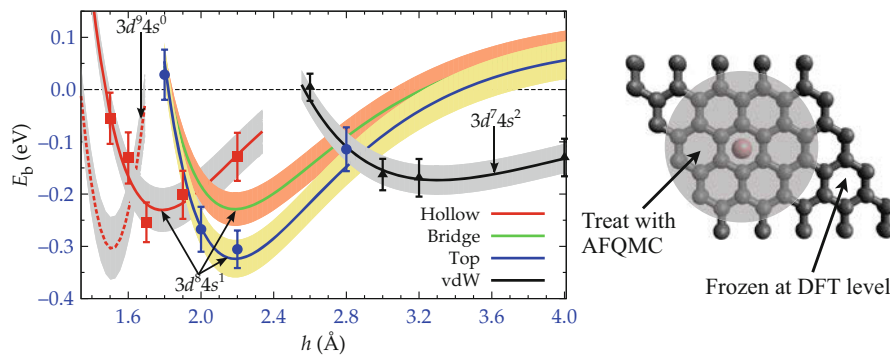


**Fig. 8** Co-adsorption on graphene and embedding AFQMC within DFT. The figure (Taken from Virgus et al. 2014) shows computed binding energy of Co on graphene, as a function of the distance $h$ between the Co atom and the graphene plane. Squares, diamonds, circles, and triangles correspond to hollow (H, as illustrated on the right), bridge (B), and top (T) sites and the van der Waals region, respectively. The dashed line indicates the low-spin H site (open squares). Shaded areas are one-$\sigma$ estimates of uncertainties, including the statistical errors in AFQMC. The right illustrates the embedding scheme for these calculations. The "inner" region with Co and the C atoms inside the circle are treated by AFQMC, with the "outer" region providing a frozen environment

# 5 Summary and Outlook

In this chapter, we have described a general computational framework for many-body calculations which combines a field-theoretic description with stochastic sampling. The approach, referred to as auxiliary-field quantum Monte Carlo (AFQMC), is based on a stochastic superposition of DFT-like calculations. We have shown how the framework can be applied to carry out ab initio electronic structure calculations. As mentioned, some additional references for further details include a set of lecture notes (Zhang 2013) (on which some of the sections in this chapter are based), a pedagogical code for lattice models written in Matlab (Nguyen et al. 2014) and a review on molecular systems (Motta and Zhang 2018).

The AFQMC approach has been applied in both condensed matter physics and quantum chemistry. It has been implemented with both plane waves/pseudopotentials and with Gaussian basis sets. We have discussed both types of calculations, as well as a combination which uses Kohn-Sham orbitals generated from plane-wave DFT as a basis to downfold the Hamiltonian for a solid. In all of these, as well as in many applications to lattice models for strong electron correlation and for ultracold atom systems, AFQMC has shown strong promise with its scalability (with system size and with parallel computing platforms), capability (total energy computation and beyond), and accuracy.

The AFQMC method has low-polynomial (cubic) scaling with system size, by using Monte Carlo sampling to treat the exponential growth of the Hilbert space. It samples the many-body ground state by a linear combination of non-orthogonal Slater determinants. The connection with independent-electron calculations, as we have highlighted, makes it straightforward to build AFQMC as a framework on top of traditional DFT or HF calculations and take advantage of the many existing technical machineries developed over the past few decades in materials modeling.

Recent developments in the computation of atomic forces and geometry optimization, and the treatment of spin-orbit coupling and general magnetic order, are manifestations of this connection. They significantly enhance the capability of stochastic methods for electronic structure. Similarly, the formulation for superconducting Hamiltonians and for embedding AFQMC in independent-electron calculations to extend length scales will broaden the reach in materials computation.

The AFQMC is approximate, because of the constraint to control the sign/phase problem. A major focus during the development of the framework has been to systematically test (and improve) the accuracy of AFQMC. A large database has now been accumulated, thanks in part to the major many-electron benchmark initiatives recently. The accuracy that can be achieved by AFQMC with its present stage of development is such that many applications are now within reach in materials modeling. A variety of new developments are possible and currently being pursued.

The structure of the open-ended random walk, as illustrated in Fig. 3, makes AFQMC ideally suited for modern high-performance computing platforms, with

exceptional capacity for parallel scaling. The rapid growth of high-performance computing resources will thus provide a strong boost to the application of AFQMC in the study of molecules and solids.

The development of AFQMC is entering an exciting new phase. A large number of possible directions can be pursued, including many opportunities for algorithmic improvements and speedups. These will be spurred forward and stimulated by growth in applications, which we hope will in turn allow more rapid realization of a general many-body computational framework for materials.

# References

Al-Saidi WA, Zhang S, Krakauer H (2006) Auxiliary-field quantum Monte Carlo calculations of molecular systems with a Gaussian basis. J Chem Phys 124(22):224101

Al-Saidi WA, Krakauer H, Zhang S (2007) A study of H + $H_2$ and several H-bonded molecules by phaseless auxiliary-field quantum Monte Carlo with plane wave and Gaussian basis sets. J Chem Phys 126(19):194105. https://doi.org/10.1063/1.2735296

Aquilante F, De Vico L, Ferre N, Ghigo G, Malmqvist P, Neogrady P, Pedersen T, Pitonak M, Reiher M, Roos B, Serrano-Andres L, Urban M, Veryazov V, Lindh R (2010) J Comput Chem 31(1):224–247. https://doi.org/10.1002/jcc.21318. The information about affiliations in this record was updated in December 2015. The record was previously connected to the following departments: Theoretical Chemistry (S) (011001039)

Baer R, Head-Gordon M, Neuhauser D (1998) Shifted-contour auxiliary field Monte Carlo for ab initio electronic structure: straddling the sign problem. J Chem Phys 109(15):6219–6226. https://doi.org/10.1063/1.477300

Bartlett RJ, MusiałM (2007) Coupled-cluster theory in quantum chemistry. Rev Mod Phys 79(1):291. https://doi.org/10.1103/RevModPhys.79.291

Blankenbecler R, Scalapino DJ, Sugar RL (1981) Monte Carlo calculations of coupled Boson-Fermion systems. I. Phys Rev D 24:2278

Booth GH, Thom AJW, Alavi A (2009) Fermion Monte Carlo without fixed nodes: a game of life, death, and annihilation in slater determinant space. J Chem Phys 131(5):054106. https://doi.org/10.1063/1.3193710

Car R, Parrinello M (1985) Unified approach for molecular dynamics and density functional theory. Phys Rev Lett 55:2471

Carlson J, Gandolfi S, Schmidt KE, Zhang S (2011) Auxiliary-field quantum Monte Carlo method for strongly paired fermions. Phys Rev A 84:061602. https://doi.org/10.1103/PhysRevA.84.061602

Ceperley DM (1995) Path integrals in the theory of condensed helium. Rev Mod Phys 67:279, and references therein

Crawford TD, Schaefer HF III (2000) An introduction to coupled cluster theory for computational chemists. Rev Comput Chem 14:33–136

Diedrich DL, Anderson JB (1992) An accurate quantum monte carlo calculation of the barrier height for the reaction $h + h_2 \rightarrow h_2 + h$. Science 258(5083):786–788. https://doi.org/10.1126/science.258.5083.786, http://science.sciencemag.org/content/258/5083/786.full.pdf

Esler KP, Kim J, Ceperley DM, Purwanto W, Walter EJ, Krakauer H, Zhang S, Kent PRC, Hennig RG, Umrigar C, Bajdich M, Kolorenc J, Mitas L, Srinivasan A (2008) Quantum Monte Carlo algorithms for electronic structure at the petascale; the Endstation project. J Phys Conf Ser 125:012057 (15pp). http://stacks.iop.org/1742-6596/125/012057

Fahy SB, Hamann DR (1990) Positive-projection Monte Carlo simulation: a new variational approach to strongly interacting fermion systems. Phys Rev Lett 65:3437

Foulkes WMC, Mitas L, Needs RJ, Rajagopal G (2001) Quantum Monte Carlo simulations of solids. Rev Mod Phys 73:33, and references therein

Hamann DR, Fahy SB (1990) Energy measurement in auxiliary-field many-electron calculations. Phys Rev B 41(16):11352

Kalos MH, Whitlock PA (1986) Monte Carlo methods, vol I. Wiley, New York

Kalos MH, Levesque D, Verlet L (1974) Helium at zero temperature with hard-sphere and other forces. Phys Rev A 9:2178

Koch H, de Merás AS, Pedersen TB (2003) Reduced scaling in electronic structure calculations using Cholesky decompositions. J Chem Phys 118(21):9481–9484. https://doi.org/10.1063/1.1578621

Kohn W (1999) Nobel lecture: Electronic structure of matter – wave functions and density functionals. Rev Mod Phys 71:1253, and references therein

LeBlanc JPF, Antipov AE, Becca F, Bulik IW, Chan GKL, Chung CM, Deng Y, Ferrero M, Henderson TM, Jiménez-Hoyos CA, Kozik E, Liu XW, Millis AJ, Prokof'ev NV, Qin M, Scuseria GE, Shi H, Svistunov BV, Tocchio LF, Tupitsyn IS, White SR, Zhang S, Zheng BX, Zhu Z, Gull E (2015) Solutions of the two-dimensional hubbard model: benchmarks and results from a wide range of numerical algorithms. Phys Rev X 5:041041. https://doi.org/10.1103/PhysRevX.5.041041

Loh EY Jr, Gubernatis JE, Scalettar RT, White SR, Scalapino DJ, Sugar R (1990) Sign problem in the numerical simulation of many-electron systems. Phys Rev B 41:9301

Ma F, Zhang S, Krakauer H (2013) Excited state calculations in solids by auxiliary-field quantum Monte Carlo. New J 15:093017. https://doi.org/10.1088/1367-2630/15/9/093017

Ma F, Purwanto W, Zhang S, Krakauer H (2015) Quantum Monte Carlo calculations in solids with downfolded hamiltonians. Phys Rev Lett 114:226401. https://doi.org/10.1103/PhysRevLett.114.226401

Ma F, Zhang S, Krakauer H (2017) Auxiliary-field quantum Monte Carlo calculations with multiple-projector pseudopotentials. Phys Rev B 95:165103. https://doi.org/10.1103/PhysRevB.95.165103

Martin RM (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, Cambridge

Moskowitz JW, Schmidt KE, Lee MA, Kalos MH (1982) A new look at correlation energy in atomic and molecular systems. II. The application of the Green's function Monte Carlo method to LiH. J Chem Phys 77:349

Motta M, Zhang S (2017) Computation of ground-state properties in molecular systems: back-propagation with auxiliary-field quantum Monte Carlo. J Chem Theory Comput 13(11):5367–5378. https://doi.org/10.1021/acs.jctc.7b00730, PMID:29053270

Motta M, Zhang S (2018, in press) Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. WIREs Comput Mol Sci. https://doi.org/10.1002/wcms.1364

Motta M, Ceperley DM, Chan GKL, Gomez JA, Gull E, Guo S, Jiménez-Hoyos CA, Lan TN, Li J, Ma F, Millis AJ, Prokof'ev NV, Ray U, Scuseria GE, Sorella S, Stoudenmire EM, Sun Q, Tupitsyn IS, White SR, Zgid D, Zhang S (2017) Towards the solution of the many-electron

problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods. Phys Rev X 7:031059. https://doi.org/10.1103/PhysRevX.7.031059

Negele JW, Orland H (1998) Quantum many-particle systems. Advanced book classics. Perseus Books, Reading

Nguyen H, Shi H, Xu J, Zhang S (2014) CPMC-lab: a matlab package for constrained path Monte Carlo calculations. Comput Phys Commun 185(12):3344–3357. https://doi.org/10.1016/j.cpc.2014.08.003, http://www.sciencedirect.com/science/article/pii/S0010465514002707

Purwanto W, Zhang S (2004) Quantum Monte Carlo method for the ground state of many-boson systems. Phys Rev E 70:056702

Purwanto W, Zhang S (2005) Correlation effects in the ground state of trapped atomic bose gases. Phys Rev A 72(5):053610

Purwanto W, Krakauer H, Zhang S (2009a) Pressure-induced diamond to $\beta$ -tin transition in bulk silicon: a quantum Monte Carlo study. Phys Rev B 80(21):214116. https://doi.org/10.1103/PhysRevB.80.214116

Purwanto W, Zhang S, Krakauer H (2009b) Excited state calculations using phaseless auxiliary-field quantum Monte Carlo: potential energy curves of low-lying $C_2$ singlet states. J Chem Phys 130(9):094107. https://doi.org/10.1063/1.3077920

Purwanto W, Krakauer H, Virgus Y, Zhang S (2011) Assessing weak hydrogen binding on $Ca^+$ centers: an accurate many-body study with large basis sets. J Chem Phys 135:164105

Purwanto W, Zhang S, Krakauer H (2013) Frozen-orbital and downfolding calculations with auxiliary-field quantum Monte Carlo. J Chem Theory Comput. https://doi.org/10.1021/ct4006486

Qin M, Shi H, Zhang S (2016) Coupling quantum Monte Carlo and independent-particle calculations: self-consistent constraint for the sign problem based on the density or the density matrix. Phys Rev B 94:235119. https://doi.org/10.1103/PhysRevB.94.235119

Rosenberg P, Shi H, Zhang S (2017) Accurate computations of Rashba spin-orbit coupling in interacting systems: from the Fermi gas to real materials. J Phys Chem Solids. https://doi.org/10.1016/j.jpcs.2017.12.026, 1710.00887

Schmidt KE, Kalos MH (1984) Few- and many-Fermion problems. In: Binder K (ed) Applications of the Monte Carlo method in statistical physics. Springer, Heidelberg

Shee J, Zhang S, Reichman DR, Friesner RA (2017) Chemical transformations approaching chemical accuracy via correlated sampling in auxiliary-field quantum Monte Carlo. J Chem Theory Comput 13(6):2667–2680. https://doi.org/10.1021/acs.jctc.7b00224, PMID: 28481546

Shi H, Zhang S (2013) Symmetry in auxiliary-field quantum Monte Carlo calculations. Phys Rev B 88:125132

Shi H, Zhang S (2016) Infinite variance in fermion quantum Monte Carlo calculations. Phys Rev E 93:033303. https://doi.org/10.1103/PhysRevE.93.033303

Shi H, Zhang S (2017) Many-body computations by stochastic sampling in Hartree-Fock-Bogoliubov space. Phys Rev B 95:045144. https://doi.org/10.1103/PhysRevB.95.045144

Sorella S, Baroni S, Car R, Parrinello M (1989) A novel technique for the simulation of interacting fermion systems. Europhys Lett 8:663

Suewattana M, Purwanto W, Zhang S, Krakauer H, Walter EJ (2007) Phaseless auxiliary-field quantum Monte Carlo calculations with plane waves and pseudopotentials: applications to atoms and molecules. Phys Rev B (Condensed Matter Mater Phys) 75(24):245123. https://doi.org/10.1103/PhysRevB.75.245123

Sugiyama G, Koonin SE (1986) Auxiliary field Monte-Carlo for quantum many-body ground states. Ann Phys (NY) 168:1

Szabo A, Ostlund N (1989) Modern quantum chemistry. McGraw-Hill, New York

Umrigar CJ, Nightingale MP, Runge KJ (1993) A diffusion Monte Carlo algorithm with very small time-step errors. J Chem Phys 99(4):2865

Virgus Y, Purwanto W, Krakauer H, Zhang S (2014) Stability, energetics, and magnetic states of cobalt adatoms on graphene. Phys Rev Lett 113:175502. https://doi.org/10.1103/PhysRevLett.113.175502

Vitali E, Shi H, Qin M, Zhang S (2016) Computation of dynamical correlation functions for many-fermion systems with auxiliary-field quantum Monte Carlo. Phys Rev B 94:085140. https://doi.org/10.1103/PhysRevB.94.085140

Wei ZC, Wu C, Li Y, Zhang S, Xiang T (2016) Majorana positivity and the fermion sign problem of quantum Monte Carlo simulations. Phys Rev Lett 116:250601. https://doi.org/10.1103/PhysRevLett.116.250601

White SR, Scalapino DJ, Sugar RL, Loh EY, Gubernatis JE, Scalettar RT (1989) Numerical study of the two-dimensional Hubbard model. Phys Rev B 40(1):506

Zhang S (1999a) Constrained path Monte Carlo for fermions. In: Nightingale MP, Umrigar CJ (eds) Quantum Monte Carlo methods in physics and chemistry. Kluwer Academic Publishers, Dordrech, cond-mat/9909090

Zhang S (1999b) Finite-temperature Monte Carlo calculations for systems with fermions. Phys Rev Lett 83:2777

Zhang S (2003) Quantum Monte Carlo methods for strongly correlated fermions. In: Sénéchal D, Tremblay AM, Bourbonnais C (eds) Theoretical methods for strongly correlated electrons. CRM series in mathematical physics, and references therein. Springer, New York

Zhang S (2013) Auxiliary-Field quantum monte carlo for correlated electron systems. In: Pavarini E, Koch E, Schollwöck U (eds) Emergent phenomena in correlated matter: modeling and simulation, vol 3. Verlag des Forschungszentrum Jülich, Jülich

Zhang S, Ceperley DM (2008) Hartree-Fock ground state of the three-dimensional electron gas. Phys Rev Lett 100:236404

Zhang S, Kalos MH (1991) Exact Monte Carlo calculations for few-electron systems. Phys Rev Lett 67:3074

Zhang S, Krakauer H (2003) Quantum Monte Carlo method using phase-free random walks with Slater determinants. Phys Rev Lett 90:136401

Zhang S, Carlson J, Gubernatis JE (1997) Constrained path Monte Carlo method for fermion ground states. Phys Rev B 55:7464

Zhang S, Krakauer H, Al-Saidi WA, Suewattana M (2005) Quantum simulations of realistic systems by auxiliary fields. Comput Phys Commun 169:394

Zheng BX, Chung CM, Corboz P, Ehlers G, Qin MP, Noack RM, Shi H, White SR, Zhang S, Chan GKL (2017) Stripe order in the underdoped region of the two-dimensional Hubbard model. Science 358(6367):1155–1160. https://doi.org/10.1126/science.aam7127

# Electrical Polarization and Orbital Magnetization: The Position Operator Tamed

**8**

Raffaele Resta

## Contents

R. Resta (✉)
Consiglio Nazionale delle Ricerche (CNR), Istituto Officina dei Materiali (IOM), Trieste, Italy

Donostia International Physics Center, San Sebastián, Spain
e-mail: resta@democritos.it

**Abstract**

Macroscopic polarization **P** and magnetization **M** are the most fundamental
concepts in textbook treatments of condensed media. They are intensive vector
quantities that intuitively carry the meaning of dipole per unit volume. But for
many years, both **P** and the orbital term in **M** evaded even a precise microscopic
definition and severely challenged quantum mechanical calculations. Contrary
to a widespread incorrect belief, **P** *has nothing to do* with the periodic charge
distribution in the bulk of a polarized crystal; analogously, the orbital term
in **M** has nothing to do with the bulk current distribution. When a bounded
sample is addressed, **P** and **M** can indeed be expressed in terms of charge
and current distributions, but the boundary contributions are essential. The
field has undergone a genuine revolution since the early 1990s. The modern
theory of polarization, based on a Berry phase, is a mature topic since the
late 1990s; it is now implemented in most first-principle electronic structure
codes. Many calculations have addressed various phenomena (ferroelectricity,
piezoelectricity, lattice dynamics, infrared spectra of liquid, and amorphous
systems) in several materials and are in spectacular agreement with experiments;
they have provided thorough understanding of the behavior of ferroelectric and
piezoelectric materials. The modern theory of orbital magnetization started in
2005, but some fundamental issues are still in development at the time of writing
(2017). Only a few first-principle calculations have appeared so far.

# 1    Introduction

The elementary definitions of macroscopic polarization **P** and orbital magnetization
**M** address a bounded sample in the large-sample limit:

$$\mathbf{P} = \frac{\mathbf{d}}{V} = \frac{1}{V} \int d\mathbf{r}\ \mathbf{r} \rho^{(\text{micro})}(\mathbf{r}) \tag{1}$$

$$\mathbf{M} = \frac{\mathbf{m}}{V} = \frac{1}{2cV} \int d\mathbf{r}\ \mathbf{r} \times \mathbf{j}^{(\text{micro})}(\mathbf{r}). \tag{2}$$

Here and in the following, we indicate with **M** the orbital term only; $\rho^{(\text{micro})}(\mathbf{r})$ and
$\mathbf{j}^{(\text{micro})}(\mathbf{r})$ are the microscopic charge and current densities, and $V$ is the sample
volume. The previous expressions are clearly dominated by surface contributions,
while instead phenomenologically **P** and **M** are bulk properties.

In general, electronic structure theory addresses bulk material properties by
adopting periodic Born-von Kármán boundary conditions (PBCs), whose main
virtue is that the system has no surface by construction. But the textbook definitions
of **P** and **M**, Eqs. (1) and (2), cannot be implemented within PBCs for a common
reason, i.e., the unboundedness of the position operator **r**. Indeed, the multiplicative
operator **r** is a "forbidden" operator within PBCs (Resta 1998).

The drawback was overcome – and the nasty position operator was effectively tamed – in the early 1990s for **P** and since 2005 onward for **M**. It was realized that for an unbounded periodic system **P** cannot be determined – even in principle – from a knowledge of the microscopic charge density $\rho^{(\mathrm{micro})}(\mathbf{r})$; analogously, **M** cannot be determined from a knowledge of the microscopic current density $\mathbf{j}^{(\mathrm{micro})}(\mathbf{r})$. The expressions for **P** and **M** in use in modern electronic structure theory (Resta 1994; Vanderbilt and King-Smith 1993; Resta and Vanderbilt 2007; Resta 2010; Thonhauser 2011; Spaldin 2012) look somewhat exotic and do not bear any resemblance whatsoever to Eqs. (1) and (2); these expressions are instead deeply rooted in the geometry of the electronic ground state.

The modern theory of polarization is a mature topic since the late 1990s and is implemented in the most popular electronic structure codes, within both density-functional and Hartree-Fock frameworks. The physical properties addressed, for real materials, are spontaneous polarization, Born (alias infrared) effective charges, infrared spectra, and piezoelectric coefficients.

The total magnetization of a sample is the sum of a spin and an orbital contribution: the latter is typically a few percent of the total. The spin term does not present a challenge for electronic structure calculations, while instead before the advent of the modern theory (2005–2006), no formula was available for evaluating orbital magnetization in bulk materials. The theory has been implemented at the first-principle level, e.g., to evaluate **M** in the ferromagnetic metals Ni, Co, and Fe (Ceresoli et al. 2010a; Lopez et al. 2012); it has also been used to address the NMR (nuclear magnetic resonance) shielding tensors in some paradigmatic cases (Thonhauser et al. 2009; Ceresoli et al. 2010b).

At variance with polarization theory, the modern theory of magnetization is partly a work in progress at the time of writing (2017); outstanding contributions appeared very recently (Bianco and Resta 2013; Marrazzo and Resta 2016; Resta 2018). These recent papers emphasize some very important features – detailed below in this chapter – which make **P** to behave very differently from **M**. This is remarkable, given the very similar textbooks definitions (Eqs. (1) and (2)) and given the common drawback therein. It is also remarkable given that the modern expressions for **P** and **M**, as integrals in reciprocal space in terms of Bloch orbitals, look apparently – and only apparently – very similar.

At the phenomenological level, there is another remarkable difference between **P** and **M** which is worth stressing at this point. **P** is well defined only for insulators, which strictly speaking exist only at zero temperature; **M** instead is well defined for both insulators and metals, even at finite temperature. Here we provide only the zero-temperature theory for **M**, as implemented in electronic structure codes (so far density-functional theory only).

The modern theories provide the values of **P** and **M** in vanishing macroscopic fields (electric and magnetic): this concept is illustrated in Sect. 2. Following it, Sect. 3 provides an overview of polarization theory, and Sect. 4 provides the theory in one dimension, where the formulation can be made particularly clear. Section 5 presents the full polarization theory, including the formulas actually implemented in first-principle calculations. Section 6 is an overview of the modern theory of

orbital magnetization, and Sect. 7 is the detailed presentation of it. Both theories are presented here in an original way, which differs from the previous published reviews (Resta 1994; Vanderbilt and Resta 2006; Resta and Vanderbilt 2007; Resta 2010; Thonhauser 2011; Spaldin 2012), and does not follow the historical developments.

## 2    Electric and Magnetic Fields

### 2.1    The Fields E and B

The modern theories – illustrated throughout the present chapter – yield the **P** and **M** values in *absence* of macroscopic fields **E** and **B**. In this section we explain with some detail what this means in the framework of modern electronic structure theory, where one addresses bulk quantities without dealing with real bounded samples.

The microscopic fields $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{B}^{(\mathrm{micro})}(\mathbf{r})$ are ideally measurable inside the material, with no reference to what happens outside a bounded sample. Their macroscopic averages **E** and **B**, i.e., the internal (or screened) macroscopic fields, are therefore the variables of choice for a first-principle description. It must be realized that, insofar as we address an infinite system with no boundaries, the macroscopic field (either **E** or **B**) is just an arbitrary boundary condition. To realize this, it is enough to focus on the electrical case for a crystalline material. The microscopic charge density is neutral in average and lattice periodical; the value of **E** is just an arbitrary boundary condition for the integration of Poisson's equation. The usual choice (performed within all electronic structure codes) is to impose a lattice-periodical Coulomb potential, i.e., $\mathbf{E} = 0$. Imposing a given nonzero value of **E** is equally legitimate (in insulators), although technically more difficult (Souza et al. 2002; Umari and Pasquarello 2002).

More generally, even for a noncrystalline and/or correlated large and macroscopically homogeneous sample, the adoption of PBCs amounts to assuming that both the Coulomb potential and the vector potential obey PBCs, ergo $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{B}^{(\mathrm{micro})}(\mathbf{r})$ average to zero over the sample.

### 2.2    The Fields D and H

Macroscopic descriptions of homogeneous media customarily adopt the electro-magnetic free-energy density $\mathscr{F}(\mathbf{E}, \mathbf{H})$, where the independent variables are the fields **E** and **H** (not **B**). The conjugate variables are

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = -4\pi\frac{\partial\mathscr{F}}{\partial\mathbf{E}}$$

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M} = -4\pi\frac{\partial\mathscr{F}}{\partial\mathbf{H}}. \tag{3}$$

The most general expansion of $\mathscr{F}(\mathbf{E}, \mathbf{H})$ up to second order around $\mathbf{E} = 0$ and $\mathbf{H} = 0$ reads

$$\mathscr{F}(\mathbf{E}, \mathbf{H}) = \mathscr{F}_0 - \mathbf{P}_0 \cdot \mathbf{E} - \mathbf{M}_0 \cdot \mathbf{H} - \frac{1}{8\pi} \mathbf{E} \overset{\leftrightarrow}{\varepsilon} \mathbf{E} - \frac{1}{8\pi} \mathbf{H} \overset{\leftrightarrow}{\mu} \mathbf{H} - \frac{1}{4\pi} \mathbf{E} \overset{\leftrightarrow}{\alpha} \mathbf{H}, \quad (4)$$

where textbooks ignore the last term.

It was discovered by Dzyaloshinskii (1960) that the tensor $\overset{\leftrightarrow}{\alpha}$ is in general nonzero in crystals where both inversion symmetry and time-reversal symmetry are absent. In fact $\mathbf{E}$ is odd under inversion and even under time-reversal, while the opposite happens for $\mathbf{H}$: therefore their combination in the last term of Eq. (4) is odd under each of the two transformation. The tensor $\overset{\leftrightarrow}{\alpha}$ is responsible for the linear magnetoelectric effect, which follows from Eq. (3): an $\mathbf{H}$ field induces polarization (at $\mathbf{E} = 0$), and conversely an $\mathbf{E}$ field induces magnetization (at $\mathbf{H} = 0$). We are not discussing magnetoelectric materials in this chapter, and we therefore set $\overset{\leftrightarrow}{\alpha} = 0$ in the following.

As for the remaining second order terms, $\overset{\leftrightarrow}{\varepsilon}$ is the macroscopic dielectric tensor, and $\overset{\leftrightarrow}{\mu}$ is the magnetic permeability tensor. We draw attention to the *minus* signs, explained, e.g., in the Landau textbook (Landau and Lifshitz 1984).

The first order terms yield the spontaneous polarization $\mathbf{P}_0 = -\partial \mathscr{F}/\partial \mathbf{E}$ and magnetization $\mathbf{M}_0 = -\partial \mathscr{F}/\partial \mathbf{H}$, where the derivatives are evaluated at $\mathbf{E} = 0$ and $\mathbf{H} = 0$.

## 2.3   B vs. H

As said above, $\mathbf{E}$ and $\mathbf{H}$ are customarily chosen as the independent variables. As for the electric field, this is clearly the natural choice. We have stressed that $\mathbf{E}$ (not $\mathbf{D}$) is the *internal* field, which is measurable in principle inside the material. The macroscopic field $\mathbf{E}$ is also a control parameter in first-principle calculations. Matters are different in the magnetic case: the *internal* field, measurable in principle inside the material, is $\mathbf{B}$, not $\mathbf{H}$. In fact the modern theory of magnetization, as discussed in the present chapter, addresses magnetization in zero $\mathbf{B}$ field.

So, why instead one adopts $\mathbf{H}$, and not $\mathbf{B}$, as the independent variable in the free-energy density? There are several reasons. Phenomenologically, the experimenter directly controls $\mathbf{E}$ (e.g., via capacitors) and $\mathbf{H}$ (via, e.g., solenoids or generally currents); the fields $\mathbf{E}$ and $\mathbf{H}$ are directly read on the instruments. Several textbooks even call $\mathbf{H}$ the "magnetic field," which I find strongly misleading. I adopt the nomenclature of the good textbooks, such as Feynman et al. (1964) and Griffiths (1999) popular textbooks, where $\mathbf{B}$ is called "magnetic field" and $\mathbf{H}$ is just "$\mathbf{H}$." All formulas have a pretty symmetric expression in terms of $\mathbf{E}$ and $\mathbf{H}$; Feynman warns however that "although the equations are analogous, the physics is not analogous." The equations would look asymmetrical and ugly using the genuine magnetic field $\mathbf{B}$ instead.

As said above, custom dictates that spontaneous magnetization is defined as $\mathbf{M}_0 = -\partial \mathscr{F}/\partial \mathbf{H}$: this is not what is addressed throughout this chapter. However $\mathbf{B}$ and $\mathbf{H}$ are related by the magnetic permeability tensor as $\mathbf{B} = \overleftrightarrow{\mu} \, \mathbf{H}$, and in normal materials the permeability difference from one (in Gaussian units) is of the order of $(1/137)^2$. There is then little difference between $\mathbf{M}_0$ as addressed in this chapter and $\mathbf{M}_0$ as customarily addressed. Whenever needed, the conversion of the response tensors from their $(\mathbf{E}, \mathbf{H})$ definition to the $(\mathbf{E}, \mathbf{B})$ counterpart requires only straightforward algebra.

## 2.4    Shape Issues

As explained so far, there is no need of addressing bounded samples and external vs. internal fields from a theoretician's viewpoint. Nonetheless a brief digression is in order, given that experiments *are* performed over bounded samples, often in external fields. Suppose a bounded macroscopic sample is inserted in a constant external field $\mathbf{E}^{(\text{ext})}$: the microscopic field $\mathbf{E}^{(\text{micro})}(\mathbf{r})$ coincides with $\mathbf{E}^{(\text{ext})}$ far away from the sample, while it is different inside because of screening effects. If we choose an homogeneous sample of *ellipsoidal shape*, then the macroscopic average of $\mathbf{E}^{(\text{micro})}(\mathbf{r})$, i.e., the macroscopic screened field $\mathbf{E}$, is constant in the bulk of the sample. The shape effects are embedded in the depolarization coefficients (Landau and Lifshitz 1984): the simplest case is the extremely oblate ellipsoid, i.e., a slab of a macroscopically homogeneous dielectric; more details are given in Resta (2010). For the slab geometry in a vanishing external field $\mathbf{E}^{(\text{ext})}$, the internal field $\mathbf{E}$ vanishes when $\mathbf{P}$ is parallel to the slab (transverse polarization), while $\mathbf{E} = -4\pi\mathbf{P}$ is the depolarization field when $\mathbf{P}$ is normal to the slab (longitudinal polarization): see Fig. 1.

The magnetic case can be discussed along similar lines. Suppose a bounded macroscopic sample is inserted in a constant external field $\mathbf{B}^{(\text{ext})}$: the microscopic field $\mathbf{B}^{(\text{micro})}(\mathbf{r})$ coincides with $\mathbf{B}^{(\text{ext})}$ far away from the sample, while it is different inside because of screening effects. If we choose an homogeneous sample of
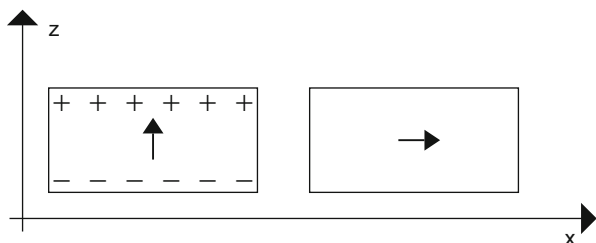


**Fig. 1** Macroscopic polarization $\mathbf{P}$ in a slab normal to $z$, for a vanishing external field $\mathbf{E}^{(\text{ext})}$. Left: when $\mathbf{P}$ is normal to the slab, a depolarizing field $\mathbf{E} = -4\pi\mathbf{P}$ is present inside the slab, and charges are present at its surface, with areal density $\sigma_{\text{surface}} = \mathbf{P} \cdot \mathbf{n}$ Right: when $\mathbf{P}$ is parallel to the slab, no depolarizing field and no surface charge is present
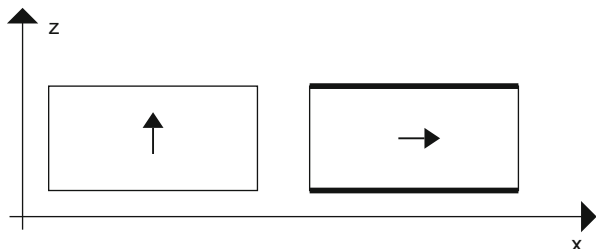
**Fig. 2** Macroscopic magnetization **M** in a slab normal to $z$, for a vanishing external field $\mathbf{B}^{(\text{ext})}$. Left: when **M** is normal to the slab, no demagnetizing field and no surface current is present. Right: when **M** is parallel to the slab, a demagnetizing field $\mathbf{B} = 4\pi\mathbf{M}$ is present inside the slab, and dissipationless currents $\mathbf{K}_{\text{surface}} = c\,\mathbf{M} \times \mathbf{n}$ flow at the surfaces

*ellipsoidal shape*, then the macroscopic average of $\mathbf{B}^{(\text{micro})}(\mathbf{r})$, i.e., the macroscopic screened field **B**, is constant in the bulk of the sample. The shape effects are embedded in the demagnetization coefficients (Landau and Lifshitz 1984). For the slab geometry in a vanishing external field $\mathbf{B}^{(\text{ext})}$, the internal field **B** vanishes when **M** is normal to the slab (longitudinal polarization), while $\mathbf{B} = -4\pi\mathbf{M}$ is the demagnetization field when **M** is parallel to the slab (transverse polarization): see Fig. 2. Notice that this is *the opposite* of what happens in the electrical case (Fig. 1).

## 3    Overview of Polarization Theory

Novel ideas about macroscopic polarization emerged in the early 1990s (Posternak et al. 1990; Resta 1992); these led to the modern theory, based on a Berry phase, which was founded by King-Smith and Vanderbilt soon afterward (King and Vanderbilt 1993). At its foundation, the modern theory was limited to a crystalline system in an independent-electron framework (either Kohn-Sham or Hartree-Fock). Later, the theory was extended to correlated and/or disordered systems (Ortiz and Martin 1994; Resta 1998).

Addressing polarization *differences* instead of polarization "itself" was instrumental to the first developments of the theory. Soon after the main breakthrough, a key paper by Vanderbilt and King-Smith (1993) reformulated the theory in a way where even polarization itself is a well-defined bulk property, although an apparently exotic one, detailed below. For the time being, let us anticipate that bulk polarization **P** is *not* a single-valued vector, at variance with its treatment in elementary textbooks.

The first calculation ever of spontaneous polarization was published by Posternak et al. (1990). The case study was BeO: it has the simplest structure where a bulk vector property is symmetry-allowed (i.e., wurtzite), and furthermore its constituents are first-row atoms. The idea was to address the macroscopic polarization of a slab of finite thickness, with faces normal to the $c$ axis, embedding it in an ad
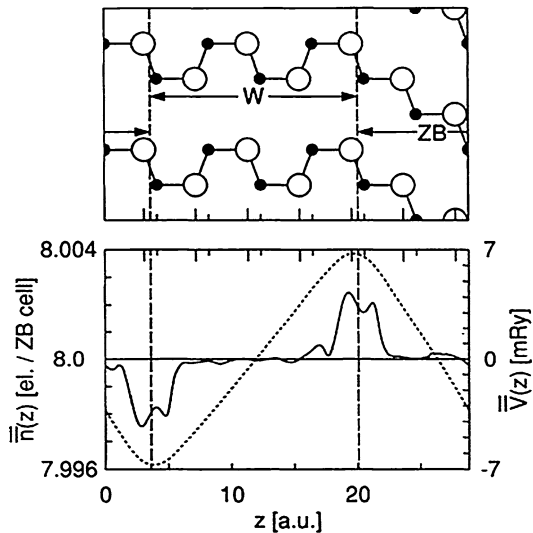
hoc medium which (i) has no bulk polarization for symmetry reasons (caveat: this was common wisdom at the time, no longer nowadays!) and (ii) does not produce any geometrical or chemical perturbation at the interface. The optimal choice is a fictitious BeO in the zincblende structure. Because of obvious computational reasons, the system is periodically replicated in a supercell geometry (Fig. 3, top panel). The self-consistent calculation shows well-localized interface charges, of opposite sign, and equal magnitude at the two nonequivalent interfaces (Fig. 3, bottom panel). The interface charge is related to the difference in polarization between the two materials: $\sigma_{\text{interface}} = \Delta\mathbf{P} \cdot \mathbf{n}$. The computer experiment provides the value of $\sigma_{\text{interface}}$, and since $\mathbf{P}$ vanishes by symmetry in the zincblende slab, one thus obtains the bulk value of $\mathbf{P}$ in the wurtzite material.

Notice that here $\mathbf{P}$ is a *longitudinal* polarization, in a depolarizing field, as in Fig. 1. Longitudinal $\mathbf{P}$ is related to the zero-field $\mathbf{P}_0$, as defined in Eq. (4), by $\mathbf{P} = \mathbf{P}_0/\varepsilon$, where $\varepsilon$ is the dielectric constant of the material.

It must be emphasized that the quantity really "measured" in this computer experiment is $\Delta\mathbf{P}$, *not* the polarization $\mathbf{P}$ itself. After the paper was published, a study of the experimental literature showed that – contrary to an incorrect widespread belief – no experimental value of $\mathbf{P}$ in any wurtzite material exists: only estimates are available. This paper marks, as said above, a change of paradigm: its main message was that polarization must be *defined* by means of *differences*, and the concept of polarization "itself" should be abandoned. With hindsight, it is nowadays pretty clear that the problem exists already at the classical level: the dipole per cell of an array of classical charges is ill defined: it depends on the (arbitrary) choice of the unit cell. Most textbooks are missing this very basic fact.

The modern theory, in its original formulation, avoided addressing the "absolute" polarization of a given equilibrium state, quite in agreement with the experiments,

which invariably measure polarization *differences*. Instead, the theory addresses differences in polarization between two states of the material that can be connected by an adiabatic switching process. The time-dependent Hamiltonian is assumed to remain insulating at all times, and the polarization difference was then *defined* (Resta 1992) as the time-integrated transient macroscopic current that flows through the insulating sample during the switching process:

$$\Delta \mathbf{P} = \mathbf{P}(\Delta t) - \mathbf{P}(0) = \int_0^{\Delta t} dt \; \mathbf{j}(t). \tag{5}$$

In the adiabatic limit, $\Delta t \rightarrow \infty$ and $\mathbf{j}(t) \rightarrow 0$, while $\Delta \mathbf{P}$ stays finite. Addressing currents (instead of charges) explains the occurrence of *phases* of the wave functions (instead of square moduli) in the modern theory. Eventually the time integration in Eq. (5) will be eliminated, leading to a two-point formula involving only the initial and final states.

The two-point formula provides two values which can be interpreted as the "absolute" polarizations $\mathbf{P}(0)$ and $\mathbf{P}(\Delta t)$ of the initial and final state, respectively. One arrives therefore at defining polarization "itself," at the price – as explained in detail below – of adopting a new paradigm: polarization is *not* a single-value vector, as instead in elementary theories.

The modern theory of polarization directly addresses polarization in zero $\mathbf{E}$ field: therefore we simplify notations by identifying $\mathbf{P}_0$ with $\mathbf{P}$ in the following. The presentation of the modern theory given here does not follow the historical developments.

## 4    Polarization Theory in One Dimension

For pedagogical reasons we start illustrating the modern theory by addressing in much detail a one-dimensional (1d) system; the 1d formalism applies as well to quasi-1d systems (e.g., polymers), with just a change of notations. The three-dimensional theory will be addressed in the following section, including the details of a first-principle implementation.

### 4.1    Polarization "Itself"

We assume a system of $N$ electrons in a segment of length $L$; the thermodynamic limit ($N \rightarrow \infty$, $L \rightarrow \infty$, $N/L = $ constant) is implicit. Neglecting irrelevant spin variables, the ground-state wave function is

$$\Psi_0 = \Psi_0(x_1, x_2, \ldots x_j, \ldots x_N). \tag{6}$$

This very general form includes disordered and/or correlated systems.

According, e.g., to the popular (Kittel 2005) textbook, **P** is nonzero when ".....the *center* of positive charge does not coincide with the *center* of negative charge" (my italics). Supposing we have 1d nuclei of charge $eZ_\ell$ at sites $X_\ell$, Kittel's definition leads to

$$P_x = e\left(\frac{1}{L}\sum_\ell Z_\ell X_\ell - \frac{2}{L}\langle\Psi_0|\sum_{j=1}^{N} x_j\ |\Psi_0\rangle\right),\tag{7}$$

where the factor of two is included to ensure charge neutrality. The two terms in parenthesis are indeed the centers of positive and negative charge, respectively.

Equation (7) makes sense and is correct for a bounded system, where (i) the sum over $\ell$ is finite and (ii) $\Psi_0$ is square-integrable; it is instead complete nonsense when PBCs are adopted. In the latter case, the sum over $\ell$ is infinite; as for the wave function, PBCs mean that for any $j$

$$\Psi_0 = \Psi_0(x_1, x_2, \ldots x_j, \ldots x_N) = \Psi_0(x_1, x_2, \ldots x_j+L, \ldots x_N).\tag{8}$$

If we restrict both the sum and the integration to a segment of length $L$, the result depends on the choice of the coordinate origin.

The key observation is that within PBCs the $x_j$ coordinates are equivalent to the angles $\vartheta_j = 2\pi x_j/L$, and the centers of positive and negative charges must be obtained from phase angles. The solution is therefore to replace Eq. (7) with

$$P_x = \frac{e}{2\pi}\text{Im ln }\ e^{i\frac{2\pi}{L}\sum_\ell Z_\ell X_\ell} - \frac{e}{\pi}\text{Im ln }\ \langle\Psi_0|e^{i\frac{2\pi}{L}\sum_j x_j}|\Psi_0\rangle;\tag{9}$$

the electronic term therein is the ground-state expectation value of a unitary and periodic operator. We may equivalently write

$$P_x = \frac{e}{2\pi}\text{Im ln }\ \langle\Psi_0|e^{i\frac{2\pi}{L}\left(\sum_\ell Z_\ell X_\ell - 2\sum_j x_j\right)}|\Psi_0\rangle = \frac{e\gamma}{2\pi} = \frac{e}{2\pi}(\gamma^{(\text{nucl})} - 2\gamma^{(\text{el})}),\tag{10}$$

where the phase angle $\gamma^{(\text{el})}$ is a Berry phase of the electronic ground state in disguise; nowadays in electronic structure jargon, this phase is dubbed "single-point Berry phase." Notice that the total phase $\gamma$ in Eq. (10) is invariant by translation of the coordinate origin, as a virtue of charge-neutrality; the nuclear and electronic phases $\gamma^{(\text{nucl})}$ and $\gamma^{(\text{el})}$ are *not* separately invariant.

## 4.2    The Polarization "Quantum"

Given that any phase angle is defined modulo $2\pi$, Eq. (10) shows that polarization itself in 1d is arbitrary modulo $e$: as anticipated above bulk polarization is a 1d lattice, not a single-valued quantity: this is the outstanding message of Vanderbilt

and King-Smith (1993). The quantum arbitrariness is a real-world feature, *not* an artefact of the theory.

When PBCs are adopted, the system has no boundary by construction, and only the bulk electronic distribution is accessible. For any bounded realization of the given linear system, the "modulo" value is determined only after the termination is specified. Among the $P_x$ values allowed by the bulk, nature will choose the minimum-energy one.

We are accustomed to assume that $P_x = 0$ in centrosymmetric linear systems: this is incorrect. The phase angles $\gamma$ compatible with centrosymmetry are either zero or $\pi$: this amounts to a $\mathbb{Z}_2$ topological classification of 1d centrosymmetric insulators, which implies either $P_x = 0 \mod e$ ($\mathbb{Z}_2$-even) or $P_x = e/2 \mod e$ ($\mathbb{Z}_2$-odd). The classification is topological, in the sense that it is impossible to continuously deform the Hamiltonian (and the ground state) of a $\mathbb{Z}_2$-even 1d system into a $\mathbb{Z}_2$-odd one while keeping the system centrosymmetric and insulating.

The proof that the end charge in centrosymmetric linear polymers is topological can be equivalently arrived at by Wannier-function counting– in the special case of independent electrons – using Eq. (39) below (Kudin et al. 2007). Similar arguments also prove the topological nature of the soliton charge in polyacetylene, first discovered by Su et al. (1979). If we insist that we want a singlet wave function – as everywhere here – then the soliton charge can only be $\pm e$. But we may relax this condition, adopting single occupancy (instead of double) for one of the Wannier functions in the soliton region: the soliton is then neutral but carries spin $\pm 1/2$.

A perspicuous illustration of the quantum feature was provided by Kudin et al. (2007), who addressed bounded quasi-1d systems of length $L$. A test case is trans-polyacetylene, whose bulk is centrosymmetric and whose bulk $P_x$ value is 0 mod $e$ ($\mathbb{Z}_2$-even). We chose two different terminations, as illustrated in Fig. 4, and we computed the dipole divided by $L$ as a function of $L$ for large $L$ values. In both cases the whole molecule is polar: therefore at finite $L$ the two dipoles are nonzero and nonquantized. The quantization is exact in the $L \rightarrow \infty$ limit: the simulations reported by Kudin et al. (2007) perspicuously show that the asymptotic value of the dipole per unit length is either zero or $e$ for the two different terminations of the same "bulk" chain.



**Fig. 4** A centrosymmetric insulating quasi-1d "crystal" with two different termination: alternant trans-polyacetylene. In this picture the "bulk" is five-monomer long. (After Kudin et al. 2007)
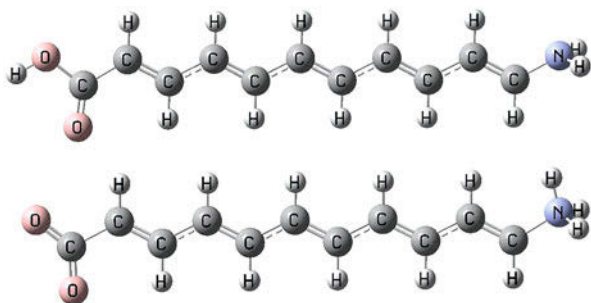
### 4.3    Polarization vs. Current

We have arrived at our main 1d result, Eq. (10), by a sort of plausibility argument. In order to prove that the $P_x$ expression given above actually is the macroscopic polarization, we have to resort to Eq. (5) and show that when $P_x$ is evaluated using the adiabatic instantaneous eigenstate $|\Psi_0\rangle$ at time $t$, its time derivative coincides with the current density. The current carried by the classical nuclei is trivial; we focus therefore on the electronic one, neglecting once more the factor of two (spinless electrons).

$$
\begin{aligned}
j_x^{(\mathrm{el})}(t) &= -\frac{e}{2\pi}\frac{d}{dt}\gamma^{(\mathrm{el})}(t) = -\frac{e}{2\pi}\frac{d}{dt}\mathrm{Im}\ln\ \langle\Psi_0(t)|e^{i\frac{2\pi}{L}\hat{x}}|\Psi_0(t)\rangle \\
&= -\frac{e}{2\pi}\mathrm{Im}\left(\frac{\langle\dot{\Psi}_0|e^{i\frac{2\pi}{L}\hat{x}}|\Psi_0\rangle}{\langle\Psi_0|e^{i\frac{2\pi}{L}\hat{x}}|\Psi_0\rangle} + \frac{\langle\Psi_0|e^{i\frac{2\pi}{L}\hat{x}}|\dot{\Psi}_0\rangle}{\langle\Psi_0|e^{i\frac{2\pi}{L}\hat{x}}|\Psi_0\rangle}\right),
\end{aligned} \tag{11}
$$

where $|\dot{\Psi}_0\rangle$ is the adiabatic time derivative of the ground eigenstate, $\hat{x} = \sum_j x_j$, and the thermodynamic limit is understood.

   We generalize the many-electron Hamiltonian as

$$
\hat{H}(\kappa) = \frac{1}{2m}\sum_{j=1}^{N}(p_j + \hbar\kappa)^2 + \hat{V}, \tag{12}
$$

where the potential $\hat{V}$ includes one- and two-body terms and is periodic of period $L$, while $\kappa$, called "flux" or "twist," is a gauge transformation. This Hamiltonian was first proposed by W. Kohn (1964), who observed that PBCs violate gauge-invariance in the conventional sense: the ground-state energy actually depends on $\kappa$. The ground eigenstate $|\Psi_0(\kappa)\rangle$, which obtains by adiabatically varying $\kappa$, has a nontrivial $\kappa$-dependence. To lowest order in $\kappa$, we get

$$
|\Psi_0(\kappa)\rangle \simeq |\Psi_0\rangle + \frac{\hbar\kappa}{m}\sum_{n\neq 0}|\Psi_n\rangle\frac{\langle\Psi_n|\ \sum_{j=1}^{N}p_j\ |\Psi_0\rangle}{E_0 - E_n}, \tag{13}
$$

where we set $|\Psi_n(0)\rangle = |\Psi_n\rangle$. We consider next the state $e^{i2\pi\hat{x}/L}|\Psi_0\rangle$: this state obeys Schrödinger equation for $\kappa = -2\pi/L$ and obeys PBCs as well. It is therefore an eigenstate of $\hat{H}(-2\pi/L)$ with energy $E_0$. If the system is an insulator, this state coincides with the ground state $|\Psi_0(-2\pi/L)\rangle$, obtained instead by adiabatically varying $\kappa$ in Eq. (12), hence:

$$
e^{i2\pi\hat{x}/L}|\Psi_0\rangle \simeq |\Psi_0\rangle - \frac{h}{mL}\sum_{n\neq 0}|\Psi_n\rangle\frac{\langle\Psi_n|\ \sum_{j=1}^{N}p_j\ |\Psi_0\rangle}{E_0 - E_n}. \tag{14}
$$

Inserting this into Eq. (11), one gets

$$j_x^{(el)} \simeq \frac{ie\hbar}{mL} \sum_{n \neq 0} \langle \dot{\Psi}_0 | \Psi_n \rangle \frac{\langle \Psi_n | \sum_{j=1}^{N} p_j | \Psi_0 \rangle}{E_0 - E_n} + \text{c.c.}, \tag{15}$$

where c.c. stays for complex conjugate. The $n = 0$ term is omitted from the sum since $\langle \dot{\Psi}_0 | \Psi_0 \rangle + \langle \Psi_0 | \dot{\Psi}_0 \rangle = 0$. This concludes our proof: in fact the right-hand member of Eq. (15) is the electronic current flowing through the system when the potential $\hat{V}$ is adiabatically varied, a well-known expression previously used by Niu and Thouless (1984) in demonstrating the quantization of particle transport.

An important subtlety is worth stressing. The relative phase between eigenvectors $|\Psi_0(\kappa)\rangle$ at two different $\kappa$-values is arbitrary. The choice performed in the conventional formulas of perturbation theory, as in Eqs. (13) and (14), correspond to the so-called parallel-transport gauge; other choices are equally legitimate. Given that the current is gauge-invariant, neglect of this phase factor (as we did here) is harmless.

## 4.4  The King-Smith and Vanderbilt Formula

The modern theory of polarization was originally formulated by King-Smith and Vanderbilt for a crystalline system of noninteracting electrons (in a mean-field sense). Here we proceed in reverse historical order: we are going to derive their formula as a special case of the previous more general one, i.e.

$$P_x = \frac{e}{2\pi}(\gamma^{(nucl)} - 2\gamma^{(el)}), \qquad \gamma^{(el)} = \text{Im ln } \langle \Psi_0 | e^{i\frac{2\pi}{L}\sum_j x_j} | \Psi_0 \rangle. \tag{16}$$

Suppose we have a crystalline system of lattice constant $a$, where we impose PBCs over $M$ linear cells: there are then $M$ equally spaced Bloch vectors in the reciprocal cell $[0, 2\pi/a]$:

$$k_s = \frac{2\pi}{Ma}s, \quad s = 0, 1, \ldots, M-1. \tag{17}$$

The size of the periodically repeated system is $L = Ma$. The one-body orbitals can be chosen to have the Bloch form:

$$\psi_{mk_s}(x + \tau) = e^{ik_s\tau}\psi_{mk_s}(x), \tag{18}$$

where $\tau = la$ is a lattice translation and $m$ is a band index. Throughout this chapter the Bloch orbitals are normalized to one over the unit cell.

There are $N/M$ occupied bands in the Slater determinant wave function, which we write (for spinless electrons) as

$$|\Psi_0\rangle = \frac{1}{\sqrt{M^N}} \mathsf{A} \prod_{m=1}^{N/M} \prod_{s=0}^{M-1} \psi_{mk_s}, \tag{19}$$

where $\mathsf{A}$ is the antisymmetrizer: the $1/\sqrt{M^N}$ factor in Eq. (19) owes to the different normalizations: $|\Psi_0\rangle$ is normalized over $L = Ma$, while the Bloch orbitals $|\psi_{m\mathbf{k}}\rangle$ are normalized over a crystal cell of length $a$. It is now expedient to define a new set of Bloch orbitals:

$$\tilde{\psi}_{mk_s}(x) = \mathrm{e}^{-i\frac{2\pi}{L}x} \psi_{mk_s}(x). \tag{20}$$

We then recast the electronic Berry phase $\gamma^{(\mathrm{el})}$, Eq. (16), as:

$$\gamma^{(\mathrm{el})} = -\mathrm{Im} \ln \langle \Psi_0 | \tilde{\Psi}_0 \rangle, \tag{21}$$

where $|\tilde{\Psi}_0\rangle$ is the Slater determinant of the $\tilde{\psi}$'s. According to a well-known theorem, the overlap among two determinants is equal to the determinant of the overlap matrix among the orbitals:

$$\gamma^{(\mathrm{el})} = -\mathrm{Im} \ln \frac{1}{M^N} \det \mathscr{S}, \tag{22}$$

where

$$\mathscr{S}_{sm,s'm'} = \int_0^L dx\, \psi_{mk_s}^*(x) \mathrm{e}^{-i\frac{2\pi}{L}x} \psi_{m'k_{s'}}(x) = M \int_0^a dx\, \psi_{mk_s}^*(x) \mathrm{e}^{-i\frac{2\pi}{L}x} \psi_{m'k_{s'}}(x), \tag{23}$$

and $\psi_{mk_M}(x) \equiv \psi_{mk_0}(x)$ is implicitly understood (so-called periodic gauge).

Owing to the orthogonality properties of the Bloch functions, the overlap matrix elements vanish except when $k_{s'} = k_s + 2\pi/L$, that is, $s' = s+1$. The $N \times N$ determinant can then be factorized into $M$ small determinants:

$$\frac{1}{M^N} \det \mathscr{S} = \prod_{s=0}^{M-1} \det S(k_s, k_{s+1}), \tag{24}$$

where for the small overlap matrix we use the notation

$$S_{m,m'}(k_s, k_{s+1}) = \int_0^a dx\, \psi_{mk_s}^*(x) \mathrm{e}^{-i\frac{2\pi}{L}x} \psi_{m'k_{s+1}}(x) = \langle u_{mk_s} | u_{m'k_{s+1}} \rangle, \tag{25}$$

where $|u_{mk}\rangle = \mathrm{e}^{-ikx} |\psi_{mk}\rangle$ is the periodic factor in the Bloch orbital. We finally get

$$\gamma^{(\mathrm{el})} \simeq -\mathrm{Im} \ln \prod_{s=0}^{M-1} \det S(k_s, k_{s+1}) : \tag{26}$$

this is the well-known expression of the modern theory of polarization, as implemented in many electronic structure codes. The formula was originally obtained by King-Smith and Vanderbilt by defining a (continuum) Berry phase as a line integral and then discretizing it.

For the sake of simplicity, we assume there is only one occupied band; Eqs. (25) and (26) become then

$$S(k_s, k_{s+1}) = \int_0^a dx \ \psi_{k_s}^*(x) e^{-i\frac{2\pi}{L}x} \psi_{k_{s+1}}(x) = \langle u_{k_s}|u_{k_{s+1}}\rangle \tag{27}$$

$$\gamma^{(\text{el})} \simeq -\text{Im} \ln \prod_{s=0}^{M-1} \langle u_{k_s}|u_{k_{s+1}}\rangle = -\sum_{s=0}^{M-1} \text{Im} \ln \langle u_{k_s}|u_{k_{s+1}}\rangle. \tag{28}$$

If we adopt a gauge such that $|u_k\rangle$ is a differentiable function of $k$, then

$$-\text{Im} \ln\langle u_k|u_{k+\Delta k}\rangle \simeq i\langle u_k|du_k/dk\rangle \, \Delta k. \tag{29}$$

In the large-$M$ limit, the sum converges to the integral

$$\gamma^{(\text{el})} = \int_0^{2\pi/a} i\langle u_k|du_k/dk\rangle \, dk. \tag{30}$$

This integral is a Berry phase (Zak 1989), and the integrand goes under the name of Berry connection. Since $|u_k\rangle$ is gauge-dependent (i.e., arbitrary by a $k$-dependent phase factor), the Berry connection is gauge-dependent as well; the Berry phase, instead, is gauge-invariant modulo $2\pi$. As shown in Sect. 4.3, the electronic current which flows through the one-dimensional sample during the switching process is

$$j_x^{(\text{el})}(t) = -\frac{ie}{2\pi}\frac{d}{dt}\int_0^{2\pi/a} \langle u_k|du_k/dk\rangle \, dk, \tag{31}$$

where the integrand is evaluated with the adiabatic instantaneous eigenstates. We remind that the simplified Eq. (31) holds for one simply occupied band in one dimension; we also remind that the system must be insulating at all times

As observed above, the continuum formulation of Eq. (30) requires $|u_k\rangle$ to be a differentiable function of $k$, while the discrete formulation of Eqs. (26) and (28) is more general. This is of foremost importance in numerical work: in fact, when the orbitals are obtained by numerical diagonalization, the choice of the gauge is erratic; further ambiguity arises from band ordering in the many-band case. We stress that this ambiguity is harmless in both Eqs. (26) and (28), which are *numerically* gauge-invariant in form. To this aim, it is essential that the Hamiltonian is diagonalized $M$ times, not $M + 1$, in order to enforce the periodic gauge, i.e.

$$\langle u_{mk_0}|u_{m'k_M}\rangle \equiv \langle u_{mk_0}|e^{-i2\pi x/a}|u_{m'k_0}\rangle. \tag{32}$$

# 5  Polarization Theory in Three Dimensions

Considerations based on the adiabatic current – similar to those leading to the simplified Eq. (31) – hold for a many-band system in three dimensions: we provide in this section the corresponding formalism in detail. The formulas given below are those implemented in most first-principle codes for electronic structure calculations, both density functional and Hartree-Fock. Since the Berry phase is included in the public distribution of such codes, it is impossible to keep track of the plethora of calculations which have appeared in the literature, for the most diverse materials. As of today (2018), the original 1993 paper by Vanderbilt and King-Smith scores 1,700 ISI citations; most of them are from papers which provided first-principle Berry-phase calculations for real materials.

In the early years (1993 onward), the modern theory has mostly addressed the ferroelectric perovskites, in order to compute (and predict) the spontaneous polarization of the various phases, the Born (a.k.a infrared) effective charges for lattice dynamics and the piezoelectric coefficients. The accuracy of the results clearly depends on the form of the functional adopted and on its reliability in describing the given material; in simple cases the overall accuracy can be evaluated at about 5%. A tabulation of several state-of-the-art results, for many perovskite oxides, has been published by Rabe and Ghosez (2007). In more recent years, the theory has been applied – as said above – to the most diverse materials, with an overall accuracy roughly of the same order, for large-gap materials at least. Small-gap materials may be more critical and may show a stronger dependence on the functional.

Another very successful application concerns the infrared spectra of amorphous and liquid materials, where the modern theory is implemented in the framework of ab initio (Car-Parrinello) molecular dynamics. The formalism is reviewed here in Sect. 5.4; the very first implementations were for amorphous silicon by Debernardi et al. (1997) and for liquid water by Silvestrelli et al. (1997). Water is arguably the most difficult liquid of all: it is hard to model the hydrogen bond with classical forces, and the light mass of hydrogen is the source of additional complications. For this reason the infrared spectrum of water has been computed again several times, with technically different first-principle ingredients. As an example, we quote here a thorough study of water and ice by Chen et al. (2008).

## 5.1  The King-Smith and Vanderbilt Formula

If $n_b$ is the number of occupied bands, then the King-Smith and Vanderbilt formula reads (per spin channel)

$$\mathbf{P}^{(el)} = -\frac{ie}{(2\pi)^3} \sum_{m=1}^{n_b} \int_{BZ} d\mathbf{k} \, \langle u_{m\mathbf{k}} | \nabla_{\mathbf{k}} u_{m\mathbf{k}} \rangle, \qquad (33)$$

where BZ is the Brillouin zone. The adiabatic time derivative of Eq. (33) yields the macroscopic current density due to the (spinless) electrons; the classical nuclear term must be added. The real vector field $i\langle u_{m\mathbf{k}}|\nabla_{\mathbf{k}}u_{m\mathbf{k}}\rangle$ is the Berry connection of band $m$: it is gauge-dependent, while $\mathbf{P}^{(\text{el})}$ is instead gauge-invariant, modulo a "quantum" detailed below.

The integral in Eq. (33) can be equivalently performed over a unit reciprocal cell, defined by the reciprocal vectors $\mathbf{G}_i$, $i = 1, 2, 3$; let the corresponding unit vectors of the direct lattice be $\mathbf{R}_i$, and the cell volume be $V_{\text{cell}} = |\mathbf{R}_1 \cdot \mathbf{R}_2 \times \mathbf{R}_3|$. We introduce the dimensionless variables $\xi_i$, defined as

$$\mathbf{k} = \sum_i \xi_i \mathbf{G}_i, \qquad \xi_i = \frac{1}{2\pi}\mathbf{k} \cdot \mathbf{R}_i. \tag{34}$$

It is expedient to evaluate $\mathbf{G}_i \cdot \mathbf{P}$ for each $i$, which in turn uniquely determine the three components of $\mathbf{P}^{(\text{el})}$. Performing the change of variables in Eq. (33), we get

$$\mathbf{G}_i \cdot \mathbf{P}^{(\text{el})} = -\frac{ie}{V_{\text{cell}}} \sum_{m=1}^{n_{\text{b}}} \int_0^1 d\xi_1 \int_0^1 d\xi_2 \int_0^1 d\xi_3 \ \langle u_m|\frac{\partial}{\partial\xi_i}u_m\rangle, \tag{35}$$

where $|u_m\rangle$ is now a function of the $\xi_i$. We focus now on a given $i$, say $i = 3$, and we write

$$\mathbf{G}_3 \cdot \mathbf{P}^{(\text{el})} = -\frac{e}{V_{\text{cell}}} \sum_{m=1}^{n_{\text{b}}} \int_0^1 d\xi_1 \int_0^1 d\xi_2 \ \gamma(\xi_1, \xi_2) \tag{36}$$

$$\gamma(\xi_1, \xi_2) = i \sum_{m=1}^{n_{\text{b}}} \int_0^1 d\xi_3 \ \langle u_m|\frac{\partial}{\partial\xi_3}u_m\rangle, \tag{37}$$

where $\gamma(\xi_1, \xi_2)$ is clearly a one-dimensional Berry phase in the variable $\xi_3$.

First-principle calculations of $\mathbf{P}$ for real materials are based on the discretization of Eqs. (36) and (37). The integral in $(\xi_1, \xi_2)$, Eq. (36), is discretized in the trivial way. Instead the many-band Berry phase, Eq. (37), is discretized in terms of overlap matrices according to Eq. (26). Numerical diagonalizations at different $(\xi_1, \xi_2, \xi_3)$ points provide eigenvectors which are affected by erratic phase factors; further ambiguity is due to band ordering. This arbitrariness does not affect the result, which is *numerically* gauge-invariant. Two features are essential, though: (i) the periodic gauge must be enforced, according to Eq. (32), and (ii) the computed $\gamma(\xi_1, \xi_2)$ at the mesh points must be equivalent to the discretization of a continuous function. As for the latter point, we observe that the phases at each mesh point are (independently) arbitrary by $2\pi$; one must choose instead the phases at adjacent mesh point by ensuring that their difference is much smaller than $2\pi$.

Notice that the Berry-phase formula, as in the previous equations, yields the electronic term for spinless electrons (i.e., for spin channel). When implemented

for real materials, this $\mathbf{P}^{(\text{el})}$ value has to be multiplied by two, and the classical nuclear term must be added. The Berry-phase formula is usually implemented – within a first-principle framework – in order to evaluate differences in polarization $\Delta\mathbf{P}$ between two states of the material that are connected by an adiabatic switching process, as in Eq. (5). The formula amounts to replacing the time integral in Eq. (5) with a two-point formula.

The Berry phase $\gamma(\xi_1, \xi_2)$ in Eq. (37) – either continuous or discretized – is gauge-invariant modulo $2\pi$ for any given $(\xi_1, \xi_2)$. If (and only if) we choose the gauge in such a way that $\gamma(\xi_1, \xi_2)$ is a continuous function of $(\xi_1, \xi_2)$, then $\mathbf{G}_3 \cdot \mathbf{P}^{(\text{el})}$ is arbitrary modulo $2\pi e / V_{\text{cell}}$. Finally, we notice that if $\mathbf{R}$ is the most general lattice vector, one has $\mathbf{G}_i \cdot \mathbf{R} = 2\pi n$, with $n$ integer: ergo the polarization quantum (for each spin occupancy) is $e\mathbf{R}/V_{\text{cell}}$. A moment's reflection shows that even the classical nuclear term is affected by the same ambiguity.

The original definition of $\Delta\mathbf{P}$ as the time integral of its derivative, as in Eq. (5), is not affected by any quantum ambiguity. The Berry-phase theory shows that $\Delta\mathbf{P}$ can be alternatively cast as a two-point formula, thus requiring the evaluation of two Berry phases only (of the initial and final state): the trade-off is the quantum ambiguity. Usually this is not a problem when evaluating $\Delta\mathbf{P}$, since its value is typically much smaller than the quantum $e\mathbf{R}/V_{\text{cell}}$, for crystal cells of moderate size. If there is any doubt about the quantum choice, one can add some intermediate points between the initial and final ones.

## 5.2    Formulation in Terms of Wannier Functions

The independent-particle (or mean-field) ground state is a Slater determinant of doubly occupied orbitals; any unitary transformation of the occupied states among themselves leaves the determinantal wave function invariant (apart for an irrelevant phase factor), and hence it leaves invariant any ground-state property.

For an insulating crystal, the Bloch orbitals of completely occupied bands can be transformed to localized Wannier orbitals (or functions) WFs. This is known since long time (Wannier 1937), but for many years the WFs have been mostly used as a formal tool; they became a popular topic in computational electronic structure only after the seminal work of Marzari and Vanderbilt (1997); the classical review on the topic is by Marzari et al. (2012).

The transformation of the Berry phase formula in terms of WFs provides an alternative, and perhaps more intuitive, viewpoint. The formal transformation was known since the 1950s, although the physical meaning of the formalism was not understood until the advent of the modern theory of polarization.

The unitary transformation which defines the WF $w_{m\mathbf{R}}(\mathbf{r})$, labeled by band $m$ and unit cell $\mathbf{R}$, within our normalization is

$$|w_{m\mathbf{R}}\rangle = \frac{V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} \, \mathrm{e}^{i\mathbf{k}\cdot\mathbf{R}} \, |\psi_{m\mathbf{k}}\rangle \; . \tag{38}$$

If one then defines the "Wannier centers" as $\mathbf{r}_{m\mathbf{R}} = \langle w_{m\mathbf{R}}|\mathbf{r}|w_{m\mathbf{R}}\rangle$, it is rather straightforward to prove that Eq. (33) is equivalent to

$$\mathbf{P}^{(\mathrm{el})} = -\frac{e}{V_{\mathrm{cell}}} \sum_{m=1}^{n_{\mathrm{b}}} \mathbf{r}_{m\mathbf{0}}. \tag{39}$$

This means that the electronic term (per spin channel) in the macroscopic polarization $\mathbf{P}$ is the dipole of the Wannier charge distributions in the central cell, divided by the cell volume. The nuclear term is obviously similar in form to Eq. (39); the sum of twice Eq. (39) plus the nuclear term is charge neutral.

WFs are severely gauge-dependent, since the phases of the $|\psi_{m\mathbf{k}}\rangle$ appearing in Eq. (38) can be chosen arbitrarily; further ambiguity arises from band crossings. However, the sum of their centers in Eq. (39) is gauge-invariant modulo a lattice vector. Therefore $\mathbf{P}^{(\mathrm{el})}$ in Eq. (39) is affected by the same "quantum" indeterminacy discussed above.

The modern theory, when formulated in terms of WFs, becomes much more intuitive and in a sense vindicates the venerable Clausius-Mossotti viewpoint: in fact, the charge distribution is partitioned into localized contributions, each providing an electric dipole, and these dipoles yield the electronic term in $\mathbf{P}$. However, it is clear from Eq. (38) that the *phase* of the Bloch orbitals is essential to arrive at the right partitioning. Any decomposition based on charge only is severely nonunique and does not provide in general the right $\mathbf{P}$, with the notable exception of the extreme case of molecular crystals.

In the latter case, in fact, we may consider the set of WFs centered on a given molecule; their total charge distribution coincides – in the weakly interacting limit – with the electron density of the isolated molecule (possibly in a local field). This justifies the elementary Clausius-Mossotti viewpoint. It is worth mentioning that the dipole of a polar molecule is routinely computed in a supercell geometry via the single-point Berry phase discussed below. The dipole value coincides with the one computed in the trivial way in the large supercell limit. Finite-size corrections, due to the local field (different in the two cases), can also be applied (Dabo et al. 2008).

The case of alkali halides – where the model is often phenomenologically used – deserves a different comment (Resta 2010). The electron densities of isolated ions (with or without fields) are quite different from the corresponding WFs charge distributions, for instance, because of orthogonality constraints: hence the Clausius-Mossotti model is *not* justified in its elementary form, despite contrary statements in the literature. We stress once more that $\mathbf{P}$ cannot be determined – even in principle – from a knowledge of the microscopic charge density of the polarized crystal.

## 5.3 The Single-Point Berry Phase

First-principle calculations for large systems, which simulate, e.g., liquids, are routinely performed in a supercell formalism and diagonalizing the Hamiltonian

at $\mathbf{k} = 0$ only (the $\Gamma$ point). The rationale beyond this is that the reciprocal cell of a large supercell is small, and any BZ integral can be approximately evaluated exploiting the mean-value theorem, i.e.

$$\int_{BZ} d\mathbf{k}\, f(\mathbf{k}) \simeq \frac{(2\pi)^3}{V_{cell}} f(0). \tag{40}$$

The Berry phase cannot be evaluated in this way, because the Berry connection is not gauge-invariant. The problem was solved by Resta (1998), where it was shown that even polarization can be evaluated in a single-point formalism.

Suppose we have a large cubic supercell of side $L$. Then the electronic term of polarization in the $z$ direction is (for each spin occupancy)

$$P_z^{(el)} = -\frac{e}{2\pi L^2} \gamma^{(el)} = \frac{e}{2\pi L^2} \mathrm{Im}\, \ln \det S^{(z)}, \tag{41}$$

where the overlap matrix elements are

$$S_{mm'}^{(z)} = \langle u_m | e^{-i2\pi z/L} | u_{m'} \rangle = \langle \psi_m | e^{-i2\pi z/L} | \psi_{m'} \rangle, \tag{42}$$

and $|u_m\rangle \equiv |\psi_m\rangle$ are the eigenstates at $\mathbf{k} = 0$. Eq. (41) directly follows from the one-dimensional formulation, Eqs. (26) and (32), in the single point case; alternatively, it also follows from Eqs. (36) and (37) when a single-point discretization is adopted. As usual, the nuclear contribution must be added to the electronic term.

There is nonetheless an outstanding difference between the one-dimensional and the three-dimensional cases: in the former case, the polarization quantum is $e$; in the latter case, the quantum is $e/L^2$. Since polarization is defined in the large-sample limit, the vanishing of the quantum implies that polarization "itself" is ill-defined for a three-dimensional noncrystalline sample, at variance with the one-dimensional (and quasi one-dimensional) case. Eq. (41) is very useful, though: see Sect. 5.4.

## 5.4    First-Principle Infrared Spectra

The main ingredient of the infrared spectrum is the imaginary part of the dielectric function $\varepsilon''(\omega)$. This can be obtained from the equilibrium fluctuations of the system at temperature $T$. When the fluctuations are evaluated within PBCs, the formula for a classical system is

$$\varepsilon''(\omega) = \frac{2\pi\omega}{3V k_B T} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \langle \mathbf{d}(t) \cdot \mathbf{d}(0) \rangle, \tag{43}$$

where $k_B$ is the Boltzmann constant, $\mathbf{d}$ is the instantaneous dipole of the sample, and $V$ is its volume; the brackets indicate a statistical average. The PBCs choice

goes under the name of Ewald-Kornfeld (Kornfeld 1924; Neumann 1983) in the molecular-dynamics jargon.

Since the pioneering works of Silvestrelli et al. (1997) and of Debernardi et al. (1997), first-principle infrared spectra of liquids and amorphous systems are routinely evaluated from Eq. (43), where the dipole $\mathbf{d} = V\mathbf{P}$ is provided by the single-point Berry phase, Eqs. (41) and (42) (plus the classical term), and its fluctuations are computed via a Car-Parrinello simulation on a cubic supercell of volume $V = L^3$. The formula, as given here, refers to classical nuclei and does not include quantum corrections.

The integral in Eq. (43) is discretized over small intervals $\Delta t$, and its evaluation needs polarization *differences* only. In fact at any discretized time $n\Delta t$, the polarization is

$$\mathbf{P}(n\Delta t) = \mathbf{P}(0) + [\mathbf{P}(\Delta t) - \mathbf{P}(0)] + [\mathbf{P}(2\Delta t) - \mathbf{P}(\Delta t)] + \dots$$
$$+ [\mathbf{P}(n\Delta t) - \mathbf{P}((n-1)\Delta t)]. \tag{44}$$

For a Car-Parrinello time step $\Delta t$ of usual magnitude, each of the terms in square parenthesis is much smaller than the polarization quantum $e/L^2$ at typical supercell sizes. This feature overcomes the quantum ambiguity.

## 6    Overview of Magnetization Theory

The magnetization of a homogeneous sample carries the intuitive meaning of magnetic dipole per unit volume. At the nonrelativistic (and semi-relativistic) level, it is the sum of a spin and an orbital contribution. Experimentally, magneto-mechanical experiments, based on the Einstein-de Haas effect, provide the two terms separately (Meyer and Asch 1961). The spin term is conceptually trivial: in absence of time-reversal symmetry, the electronic spin density is nonzero, and spin magnetization is proportional to the macroscopic average of the spin density. In the crystalline case, electronic structure codes routinely compute the spin density, which is a lattice-periodic function: its cell average yields the spin magnetization. Here we are concerned with orbital magnetization only, indicating it with the symbol $\mathbf{M}$.

The modern theory addresses $\mathbf{M}$ in zero macroscopic $\mathbf{B}$ field, where the time-reversal symmetry is broken by some internal microscopic field, e.g., through the spontaneous development of ferromagnetic order or via spin-orbit coupling to a background of ordered local moments; in the crystalline case, the Hamiltonian is lattice periodic.

The starting date of the modern theory is 2005; important developments are much more recent. Originally, the theory addressed crystalline systems, and $\mathbf{M}$ was expressed as a Brillouin zone integral for insulators, more generally as a Fermi-volume integral for both insulators and metals. The integrand is based on $\mathbf{k}$-derivatives of Bloch orbitals; the $\mathbf{M}$ expression looks therefore superficially kind of similar to the integrand in $\mathbf{P}$ (i.e., the Berry connection). It was early realized,

however, that – at variance with the case of **P** – the integrand is gauge-invariant, and **M** is free from any "quantum" ambiguity. The overwhelming consequences of these two features have been fully realized only since 2013 onward and will be emphasized below. For instance, tinkering with the boundary of a finite sample may alter the value of **P** (Sect. 4.2), but not of **M**. After 2013 the theory may deal with **M** in noncrystalline systems and even in inhomogeneous ones (e.g., heterojunctions). At the time of writing (2017), there is no theory for correlated ground-state wave functions. The magnetization formula for crystalline systems of noninteracting electrons was generalized to finite temperatures as well. We are not dealing with the topic here: we refer to the original literature or to a previous review by Resta (2010).

As we previously did for the case of **P**, the presentation is based on the most recent findings (Bianco and Resta 2013; Marrazzo and Resta 2016; Resta 2018) and does not follow the historical development of the theory.

## 7    Magnetization Theory

### 7.1    Ground-State Projector

In the independent-electron framework (in a mean-field sense), all ground-state properties are embedded in the one-particle density matrix $\mathscr{P}$, also called ground-state projector. For the sake of simplicity, we give once more the formulas for spinless electrons (or per spin channel).

First we address a bounded sample, where the occupied orbitals are square-integrable: this case is usually called "open boundary conditions" (OBCs). If $|\varphi_j\rangle$ are the single-particle orbitals with eigenvalues $\varepsilon_j$, the ground-state projector is

$$\mathscr{P} = \sum_{\varepsilon_j < \mu} |\varphi_j\rangle\langle\varphi_j|, \tag{45}$$

where $\mu$ is the Fermi level.

We are considering here a macroscopically homogenous system, either disordered or crystalline (in the latter case a crystallite cut from the bulk). If $O$ is any operator defining a physical property, the corresponding intensive ground-state observable is given by a trace per unit volume

$$\frac{\langle O \rangle}{V} = \text{Tr}_V\{\mathscr{P}O\} = \frac{1}{V}\int d\mathbf{r}\,\langle\mathbf{r}|\mathscr{P}O|\mathbf{r}\rangle, \tag{46}$$

in the large-$V$ limit. The electronic contribution to **P** and orbital magnetization **M** – Eqs. (1) and (2) – are simple observables in this class. For instance, magnetization is

$$\mathbf{M} = -\frac{e}{2c}\mathrm{Tr}_V\{\mathscr{P}\mathbf{r} \times \mathbf{v}\}, \tag{47}$$

where $\mathbf{v} = i[\mathscr{H}, \mathbf{r}]/\hbar$ is the velocity operator.

The orbitals, and hence the projector, depend of the choice of the gauge in the Hamiltonian $\mathscr{H}$. Once the gauge in $\mathscr{H}$ is fixed, the projector is gauge-invariant, in the sense that it is invariant by unitary transformations of the occupied orbitals among themselves; the many-electron wave function (i.e., the Slater determinant) is gauge-invariant as well.

If instead we address an unbounded crystalline sample within PBCs, the ground-state projector is cast as a BZ integral in terms of the Bloch orbitals $|\psi_{j\mathbf{k}}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{j\mathbf{k}}\rangle$; here they are normalized to one over the unit cell of volume $V_{\mathrm{cell}}$. The ground-state projector is then given by

$$\mathscr{P} = \frac{V_{\mathrm{cell}}}{(2\pi)^3}\sum_j \int_{\mathrm{BZ}} d\mathbf{k}\,\boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}})|\psi_{j\mathbf{k}}\rangle\langle\psi_{j\mathbf{k}}|, \tag{48}$$

where $\varepsilon_{j\mathbf{k}}$ are the band energies and $\boldsymbol{\theta}$ is the step function. This expression applies to both insulators ($\mu$ in a gap) and metals ($\mu$ across a band).

The PBCs projector is a lattice-periodical operator, commuting with the translation group:

$$\langle\mathbf{r}|\mathscr{P}|\mathbf{r}'\rangle = \langle\mathbf{r}+\mathbf{R}|\mathscr{P}|\mathbf{r}'+\mathbf{R}\rangle, \tag{49}$$

where $\mathbf{R}$ is a lattice vector. It is also "nearsighted" Kohn (1996), i.e.

$$\langle\mathbf{r}|\mathscr{P}|\mathbf{r}'\rangle \to 0 \quad \text{for } |\mathbf{r}-\mathbf{r}'| \to \infty; \tag{50}$$

the decay is quasi-exponential (i.e., exponential times a power) in insulators (He and Vanderbilt 2001) and power law in metals.

In the PBCs framework, an intensive observable of the electronic ground state is still expressed as a trace per unit volume, where Eq. (46) is replaced by

$$\mathrm{Tr}_V\{\mathscr{P}O\} = \frac{1}{V_{\mathrm{cell}}}\int_{\mathrm{cell}} d\mathbf{r}\,\langle\mathbf{r}|\mathscr{P}O|\mathbf{r}\rangle, \tag{51}$$

but in order for this to make sense, even $O$ must be a lattice-periodical operator, i.e.

$$\langle\mathbf{r}|O|\mathbf{r}'\rangle = \langle\mathbf{r}+\mathbf{R}|O|\mathbf{r}'+\mathbf{R}\rangle. \tag{52}$$

This clearly rules out $\mathbf{P}$ and $\mathbf{M}$, if we wish to make use of Eqs. (1) and (2), where the position operator $\mathbf{r}$ (unbounded and nonperiodical) plays a dominant role.

Finally we observe that, in both the OBCs and PBCs cases, the ground-state projector can be formally written as

$$\mathscr{P} = \boldsymbol{\theta}(\mu - \mathscr{H}). \tag{53}$$

All ground-state properties must be invariant by a translation of the energy zero (bar possibly the energy itself); Eq. (53) perspicuously shows that $\mathscr{P}$ enjoys such property. More generally, any ground-state observable is – at least in principle – a function of $\mathscr{H} - \mu$.

## 7.2    The Magnetization of a Bounded Sample

For a bounded sample, we may safely adopt the definition of Eqs. (2) and (46); it is however expedient to write the velocity operator in the equivalent form $\mathbf{v} = i[(\mathscr{H} - \mu), \mathbf{r}]/\hbar$, whence

$$\mathbf{M} = -\frac{ie}{2\hbar cV} \mathrm{Tr} \{\mathscr{P}\mathbf{r} \times (\mathscr{H} - \mu)\mathbf{r}\}, \tag{54}$$

since $\mathbf{r} \times \mathbf{r} = 0$. Insofar as the system remains finite, the appearance of $\mu$ in Eq. (54) looks irrelevant. But our choice has the virtue that $\mathbf{M}$ is *explicitly* a function of $\mathscr{H} - \mu$: this is essential in the thermodynamic limit for metals, where the spectrum becomes dense around $\mu$.

So far, we have identically transformed the textbook definition of Eq. (2) into the equivalent Eq. (54): this is still plagued by the same original drawback: the $\mathbf{r}$-space integral includes an extensive contribution from the boundary. Next we are going to tame the position operator by performing a transformation similar in spirit to an integration by parts, where the same integrated value obtains from two very different *integrands*. Eventually, the trace will be boundary-insensitive in the large-$V$ limit.

Using the cyclic properties of the trace, the Cartesian components of $\mathbf{M}$ are

$$M_\gamma = \frac{ie}{2\hbar cV} \varepsilon_{\gamma\alpha\beta} \mathrm{Tr} \{(\mathscr{H} - \mu)r_\alpha \mathscr{P}r_\beta\}. \tag{55}$$

where $\varepsilon_{\gamma\alpha\beta}$ is the antisymmetric tensor and the sum over repeated indices is implicit (here and throughout). The following lemma is then very useful. Let $O$ be any Hermitian operator commuting with $\mathscr{P}$. Then

$$\mathrm{Tr} \{O [r_\alpha, \mathscr{P}] [r_\beta, \mathscr{P}]\} = -\mathrm{Tr} \{O\mathscr{P}r_\alpha r_\beta\} - \mathrm{Tr} \{O(2\mathscr{P} - I)r_\alpha \mathscr{P}r_\beta\}. \tag{56}$$

We apply this lemma by identifying $O$ with the operator $|\mathscr{H} - \mu| = (\mathscr{H} - \mu)(I - 2\mathscr{P})$ to get

$$M_\gamma = -\frac{ie}{2\hbar cV} \varepsilon_{\gamma\alpha\beta} \mathrm{Tr} \{|\mathscr{H} - \mu| [r_\alpha, \mathscr{P}] [r_\beta, \mathscr{P}]\}, \tag{57}$$

where we have used $(I - 2\mathscr{P})^2 = I$.

If we now define the vector field

$$\mathfrak{M}(\mathbf{r}) = -\frac{ie}{2\hbar c} \langle \mathbf{r} | |\mathscr{H} - \mu| [\mathbf{r}, \mathscr{P}] \times [\mathbf{r}, \mathscr{P}] | \mathbf{r} \rangle, \tag{58}$$

the previous main result can be cast in the very compact form

$$\mathbf{M} = \frac{1}{V} \int d\mathbf{r} \, \mathfrak{M}(\mathbf{r}), \tag{59}$$

where we remind that the formula as it stands applies to spinless electrons. As proved here, the integrated value in Eq. (59) is identical to Eqs. (2) and (54) at any sample size; the key difference is that the *integrands* are quite different: Eqs. (2) and (54) are dominated by boundary contributions, while Eq. (59) is free from such drawback. Indeed, Eq. (59) is a *local* definition of magnetization: if the bounded sample is a crystallite (cut from a bulk crystal), then a definition of **M** equivalent to Eq. (59) in the large-sample limit is

$$\mathbf{M} = \frac{1}{V_{\text{cell}}} \int_{V_{\text{cell}}} d\mathbf{r} \, \mathfrak{M}(\mathbf{r}), \tag{60}$$

where $V_{\text{cell}}$ is a crystal cell in the center region of the crystallite. Obviously, a similar property *does not* hold for the original integral (Eq. (2)). Owing to the locality property, the definition of Eq. (60) can also deal – with obvious modifications – with noncrystalline or inhomogeneous samples, like heterojunctions.

In all of the previous equations, the large-sample limit is understood; it has been demonstrated (Bianco and Resta 2013; Marrazzo and Resta 2016) that the convergence of Eq. (60) is much faster than the convergence of Eqs. (2), (54), and (59). Simulations based on a model Hamiltonian in two dimensions have proved that the convergence of Eqs. (2), (54), and (59) is of the order $1/L$, where $L$ is the linear dimension of the sample, in both insulators and metals. Instead the convergence of Eq. (60) is exponential in insulators and $L^{-\alpha}$, with $\alpha > 1$, in metals. The difference between the two owes to the different "nearsightedness" of $\mathscr{P}$, mentioned above (Kohn 1996; He and Vanderbilt 2001).

### 7.3 The Magnetization of an Unbounded Crystalline Sample

The definition of the vector field $\mathfrak{M}(\mathbf{r})$, Eq. (58), applies as it stands even to an unbounded sample within PBCs. In the crystalline case, the $\mathscr{P}$ projector therein is expressed in terms of Bloch orbitals as in Eq. (48). It is easy to verify that the operator $[\mathbf{r}, \mathscr{P}]$ commutes with the lattice translations, and therefore $\mathfrak{M}(\mathbf{r})$ is a lattice-periodical function. In this sense the commutator with $\mathscr{P}$ effectively "tames" the nasty, unbounded, position operator $\mathbf{r}$. The expression for **M** as a trace per unit volume is therefore formally identical to Eq. (60), found above for a *bounded*

sample. We show next how to express it as a BZ integral in terms of Bloch orbitals; in metals the BZ integral is actually a Fermi volume integral, owing to the $\theta$ function entering the $\mathscr{P}$ definition (Eq. (48)).

We start rewriting Eq. (48) as

$$\langle \mathbf{r} | \mathscr{P} | \mathbf{r}' \rangle = \frac{V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} \ e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \langle \mathbf{r} | \mathscr{P}_{\mathbf{k}} | \mathbf{r}' \rangle \tag{61}$$

$$\mathscr{P}_{\mathbf{k}} = \sum_j \theta(\mu - \varepsilon_{j\mathbf{k}}) |u_{j\mathbf{k}}\rangle \langle u_{j\mathbf{k}}|. \tag{62}$$

The $\mathscr{P}_{\mathbf{k}}$ projector is gauge-invariant in the generalized Marzari-Vanderbilt sense (Marzari and Vanderbilt 1997; Marzari et al. 2012), i.e., it is invariant for any unitary transformation of the occupied $|u_{j\mathbf{k}}\rangle$ at the given $\mathbf{k}$; it is also periodic in $\mathbf{k}$ (even when the Chern invariant is nonzero). Therefore taking the $\mathbf{k}$-gradient inside the integral, we get

$$0 = i(\mathbf{r} - \mathbf{r}') \langle \mathbf{r} | \mathscr{P} | \mathbf{r}' \rangle + \frac{V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} \ e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \langle \mathbf{r} | \nabla_{\mathbf{k}} \mathscr{P}_{\mathbf{k}} | \mathbf{r}' \rangle; \tag{63}$$

$$[\mathbf{r}, \mathscr{P}] = \frac{i V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} \ e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \nabla_{\mathbf{k}} \mathscr{P}_{\mathbf{k}}. \tag{64}$$

The Hamiltonian $\mathscr{H}$ can also be written as a BZ integral, hence

$$|\mathscr{H} - \mu| = \frac{V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} \ e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} |\mathscr{H}_{\mathbf{k}} - \mu|, \tag{65}$$

where $\mathscr{H}_{\mathbf{k}} = e^{-i\mathbf{k} \cdot \mathbf{r}} \mathscr{H} e^{i\mathbf{k} \cdot \mathbf{r}}$.

We thus get the relevant operator identity:

$$|\mathscr{H} - \mu| [r_\alpha, \mathscr{P}] [r_\beta, \mathscr{P}] = -\frac{V_{\text{cell}}}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} \ e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} |\mathscr{H}_{\mathbf{k}} - \mu| (\partial_{k_\alpha} \mathscr{P}_{\mathbf{k}})(\partial_{k_\beta} \mathscr{P}_{\mathbf{k}}). \tag{66}$$

Notice that the three BZ integrals entering the product on the l.h.s. – from Eqs. (64) and (65) – eventually contract to a single BZ integral. This owes to the fact that both $\langle \mathbf{r} | \mathscr{H}_{\mathbf{k}} | \mathbf{r}' \rangle$ and $\langle \mathbf{r} | \nabla \mathscr{P}_{\mathbf{k}} | \mathbf{r}' \rangle$ are lattice-periodical in $\mathbf{r}$ and $\mathbf{r}'$ *separately*.

The ultimate formula (Resta 2018) for the orbital magnetization of a crystalline system is then, after Eqs. (58) and (60)

$$\mathbf{M} = \frac{ie}{2\hbar c} \int_{\text{BZ}} \frac{d\mathbf{k}}{(2\pi)^3} \text{Tr} \{ |\mathscr{H}_{\mathbf{k}} - \mu| (\nabla_{\mathbf{k}} \mathscr{P}_{\mathbf{k}}) \times (\nabla_{\mathbf{k}} \mathscr{P}_{\mathbf{k}}) \}. \tag{67}$$

## 7.4 Insulators and Metals

Our main magnetization formula (Eq. (67)) holds for both insulators and metals: the latter case requires further clarification. We start from the $\mathscr{P}_\mathbf{k}$ definition, Eq. (62), and we further adopt a gauge where $|u_{j\mathbf{k}}\rangle$ is a differentiable function of $\mathbf{k}$ (this is always possible, even for topologically nontrivial solids): then

$$
\nabla_\mathbf{k}\mathscr{P}_\mathbf{k} = \sum_j \boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}})(|\nabla_\mathbf{k}u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}| + |\nabla_\mathbf{k}u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|)
$$
$$
- \sum_j \delta(\mu - \varepsilon_{j\mathbf{k}})\nabla_\mathbf{k}\varepsilon_{j\mathbf{k}}\,|u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|. \tag{68}
$$

The second line vanishes in insulators but is singular in metals. Nonetheless the singularity does not affect Eq. (67), given that it cancels in the antisymmetrized product. We may therefore safely neglect the second line of Eq. (68) in the following. The integrand in Eq. (67) is a continuous function of $\mathbf{k}$ in insulators and piecewise continuous in metals; Eq. (67) is indeed a well-defined Fermi-volume integral in both cases. As for the first line of Eq. (68), we notice that, while $\nabla_\mathbf{k}\mathscr{P}_\mathbf{k}$ is gauge-invariant and Hermitian, the operators $|\nabla_\mathbf{k}u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|$ and $|u_{j\mathbf{k}}\rangle\langle\nabla_\mathbf{k}u_{j\mathbf{k}}|$ are in general gauge-dependent and non-Hermitian: in fact the trace of $|\nabla_\mathbf{k}u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|$, times $i$, is nothing else than the Berry connection of band $j$, i.e., the essential ingredient of polarization theory (see Sect. 5).

At this point we wish to make contact with the original formula of magnetization theory (Xiao et al. 2005; Thonhauser et al. 2005; Ceresoli et al. 2006), as reported in the previous reviews (Resta 2010; Thonhauser 2011),and implemented in a few first-principle calculations:

$$
M_\gamma = -\frac{ie}{2\hbar c}\varepsilon_{\gamma\alpha\beta}\sum_{\varepsilon_{j\mathbf{k}}<\mu}\int_{\mathrm{BZ}}d\mathbf{k}\,\langle\partial_{k_\alpha}u_{j\mathbf{k}}|(\mathscr{H}_\mathbf{k} + \varepsilon_{j\mathbf{k}} - 2\mu)|\partial_{k_\beta}u_{j\mathbf{k}}\rangle. \tag{69}
$$

It is important to observe that Eq. (69) requires the so-called Hamiltonian gauge, i.e., the $|u_{j\mathbf{k}}\rangle$ are eigenstates of $\mathscr{H}_\mathbf{k}$: unitary mixing of them – à la Marzari-Vanderbilt (Marzari and Vanderbilt 1997; Marzari et al. 2012) – is *not* permitted. Our more general Eq. (67), instead, is fully gauge-invariant.

In order to prove the equivalence, it is expedient to introduce a lemma. Let $O_\mathbf{k}$, with eigenvalues $o_{j\mathbf{k}}$, be any operator which commutes with $\mathscr{H}_\mathbf{k}$, i.e.

$$
O_\mathbf{k} = \sum_j |u_{j\mathbf{k}}\rangle o_{j\mathbf{k}}\langle u_{j\mathbf{k}}|. \tag{70}
$$

Then the following identity holds:

$$\text{Tr}\{O_{\mathbf{k}}(\partial_{k_\alpha}\mathscr{P}_{\mathbf{k}})(\partial_{k_\beta}\mathscr{P}_{\mathbf{k}})\} = \sum_j \boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}}) o_{j\mathbf{k}} \langle \partial_{k_\alpha} u_{j\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle \tag{71}$$

$$+ \sum_j \boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}}) \langle \partial_{k_\beta} u_{j\mathbf{k}} | O_{\mathbf{k}}(I - 2\mathscr{P}_{\mathbf{k}}) | \partial_{k_\alpha} u_{j\mathbf{k}} \rangle.$$

This identity is somewhat reminiscent of Eq. (56) and is proved via a straightforward although somewhat tedious calculation.

We now identify $O_{\mathbf{k}}$ with $|\mathscr{H}_{\mathbf{k}} - \mu| = (\mathscr{H}_{\mathbf{k}} - \mu)(I - 2\mathscr{P}_{\mathbf{k}})$ and we get

$$\text{Tr}\{|\mathscr{H}_{\mathbf{k}} - \mu|(\partial_{k_\alpha}\mathscr{P}_{\mathbf{k}})(\partial_{k_\beta}\mathscr{P}_{\mathbf{k}})\} = \sum_j \boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}})(\mu - \varepsilon_{j\mathbf{k}}) \langle \partial_{k_\alpha} u_{j\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle$$

$$\tag{72}$$

$$+ \sum_j \boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}}) \langle \partial_{k_\beta} u_{j\mathbf{k}} | (\mathscr{H}_{\mathbf{k}} - \mu) | \partial_{k_\alpha} u_{j\mathbf{k}} \rangle,$$

where we have exploited $(I - 2\mathscr{P}_{\mathbf{k}})^2 = I$. Antisymmetrization yields

$$\varepsilon_{\gamma\alpha\beta} \text{Tr}\{|\mathscr{H}_{\mathbf{k}} - \mu|(\partial_{k_\alpha}\mathscr{P}_{\mathbf{k}})(\partial_{k_\beta}\mathscr{P}_{\mathbf{k}})\}$$

$$= \varepsilon_{\gamma\alpha\beta} \sum_j \boldsymbol{\theta}(\mu - \varepsilon_{j\mathbf{k}}) \langle \partial_{k_\alpha} u_{j\mathbf{k}} | (2\mu - \varepsilon_{j\mathbf{k}} - \mathscr{H}_{\mathbf{k}}) | \partial_{k_\beta} u_{j\mathbf{k}} \rangle, \tag{73}$$

and substitution into Eq. (67) concludes our proof.

## 8    Conclusions

Macroscopic polarization $\mathbf{P}$ and orbital magnetization $\mathbf{M}$ are phenomenologically known as intensive ground-state properties of condensed matter. Their elementary definitions, Eqs. (1) and (2), are very similar and are affected by the same drawback: the unboundedness of the position $\mathbf{r}$ therein. Because of this feature, the elementary definitions *cannot* be adopted to address macroscopic systems: Eqs. (1) and (2) are incompatible with periodic Born-von Kármán boundary conditions, which are at the root of all condensed matter theory and which make bulk physical properties accessible to computation (for both crystalline and noncrystalline materials).

The two very important observables $\mathbf{P}$ and $\mathbf{M}$ have remained problematic for many years; the nasty position operator $\mathbf{r}$ has been effectively "tamed" in the early 1990s in the case of $\mathbf{P}$ and since 2005 onward in the case of $\mathbf{M}$. In both cases the formulas adopted by modern condensed matter physics, as presented in detail throughout this chapter, do not bear any resemblance to Eqs. (1) and (2). In the special case of a crystalline system of independent electrons, both $\mathbf{P}$ and $\mathbf{M}$ are expressed as reciprocal-space integrals, where no reference to the position operator $\mathbf{r}$ is made. The integrands therein look superficially – and only superficially – rather

similar; here we have emphasized instead the key features which make the intensive observables **P** and **M**, as defined by the modern theories, profoundly different between themselves.

The qualitative difference came out as a big surprise, given the close similarity of Eqs. (1) and (2); we may spell out such difference as follows. Macroscopic polarization **P** is a Berry phase of the electronic ground-state wave function and does not admit any local representation: there is no "polarization density" to speak of. Orbital magnetization **M** can be defined – similarly to **P** – as a reciprocal-space integral, but, at variance with **P**, it also admits a local representation with a well-defined density in coordinate space. The modern theory of magnetization may even address macroscopically inhomogeneous systems (e.g., heterojunctions), which are *not* accessible to the modern theory of polarization.

The modern theory of polarization is nowadays a mature topic. Its formalism, presented here in detail in Sect. 5, is implemented as a standard option in most density-functional-theory electronic structure codes, based on first-principle ingredients. Because of this, there have been hundreds of applications to the most diverse materials, starting with the most common ferroelectrics and piezoelectrics. The accuracy and predictive power are remarkable, of the order of 5%–10% for nonproblematic materials. An extension of the theory – associated to Car-Parrinello simulations – has also provided very accurate infrared spectra for amorphous and liquid materials, including the paradigmatic case of water.

The modern theory of orbital magnetization is instead much less popular; there are several reasons for this. Bulk magnetization is the sum of an orbital term and a spin term: the latter is the dominant one in the most common materials. In the paradigmatic ferromagnetic metals Fe, Co, and Ni, the orbital contribution to the net magnetization is between 5% and 10%. Furthermore in most materials, the orbital magnetization is due to very localized $d$ orbitals, whose orbital moment can be evaluated semiquantitatively via some crude approximation. Finally, a technical complication concerns pseudopotential implementations: the core electrons do not contribute to electrical polarization but contribute instead to orbital magnetization. Therefore core-reconstruction corrections have to be added to the formulas presented here, whenever implemented in a plane-wave pseudopotential framework.

It is expected that the modern theory of polarization, in the different variants described here, will remain a staple of electronic structure theory for years to come. As for the modern theory of magnetization, its appeal for addressing real materials from first principles is limited so far. It will be enhanced if some all-electron implementations become available and if materials where the orbital term is the dominant one are discovered.

# References

Bianco R, Resta R (2013) Orbital magnetization as a local property. Phys Rev Lett 110:087202
Ceresoli D, Thonhauser T, Vanderbilt D, Resta R (2006) Orbital magnetization in crystalline solids: multi-band insulators, Chern insulators, and metals. Phys Rev B 74:024408

Ceresoli D, Gertsmann U, Seitsonen AP, Mauri F (2010a) First-principles theory of orbital magnetization. Phys Rev B 81:060409

Ceresoli D, Marzari N, Lopez NG, Thonhauser T (2010b) Ab initio converse NMR approach for pseudopotentials. Phys Rev B 81:184424

Chen W, Sharma M, Resta R, Galli G, Car R (2008) Role of dipolar correlations in the infrared spectra of water and ice. Phys Rev B 77:245114

Dabo I, Kozinsky IB, Singh-Miller NE, Marzari N (2008) Electrostatics in periodic boundary conditions and real-space corrections. Phys Rev B 77:115139

Debernardi A, Bernasconi M, Cardona M, Parrinello M (1997) Infrared absorption in amorphous silicon from ab initio molecular dynamics. Appl Phys Lett 71:2692–2694

Dzyaloshinskii IE (1960) On the magneto-electrical effect in antiferromagnets. Sov Phys JETP 10:628–629

Feynman RP, Leighton RB, Sands M (1964) The Feynman lectures in physics. Addison Wesley, Reading. II-36-6

Griffiths DJ (1999) Introduction to electrodynamics. Prentice-Hall, New Jersey

He LX, Vanderbilt D (2001) Exponential decay properties of Wannier functions and related quantities. Phys Rev Lett 86:5341–5344

King-Smith D, Vanderbilt D (1993) Theory of polarization of crystalline solids. Phys Rev B 48:1651–1654

Kittel C (2005) Introduction to solid state physics. Wiley, Hoboken

Kohn W (1964) Theory of the insulating state. Phys Rev 133:A171–A181

Kohn W (1996) Density functional and density matrix method scaling linearly with the number of atoms. Phys Rev Lett 76:3168–3171

Kornfeld H (1924) Die Berechnung elektrostatischer Potentiale und der Energie von Dipol- und Quadrupolgittern. Z Phys 22:27–43

Kudin KN, Car R, Resta R (2007) Quantization of the dipole moment and of the end charges in push-pull polymers. J Chem Phys 127:194902

Landau LD, Lifshitz EM (1984) Electrodynamics of continuous media. Pergamon Press, Oxford

Lopez MG, Vanderbilt D, Thonhauser T, Souza I (2012) Wannier-based calculation of the orbital magnetization in crystals. Phys Rev B 85:014435

Marrazzo A, Resta R (2016) Irrelevance of the boundary on the magnetization of metals. Phys Rev Lett 116:137201

Marzari N, Vanderbilt D (1997) Maximally localized generalized Wannier functions for composite energy bands. Phys Rev B 56:12847–12865

Marzari N, Mostofi AA, Yates JR, Souza I, Vanderbilt D (2012) Maximally localized Wannier functions: theory and applications. Rev Mod Phys 84:1419–1475

Meyer AJP, Asch G (1961) Experimental g' and g values of Fe, Co, Ni, and their alloys. J Appl Phys 32:S330

Neumann M (1983) Dipole moment fluctuation formulas in computer simulations of polar systems. Molec Phys 50:841–858

Niu Q, Thouless DJ (1984) Quantised adiabatic charge transport in the presence of substrate disorder and many-body interaction. J Phys A 17:2453–2462

Ortiz G, Martin RM (1994) Macroscopic polarization as a geometric quantum phase: many-body formulation. Phys Rev B 49:14202–14210

Posternak M, Baldereschi A, Catellani A, Resta R (1990) Ab-initio study of the spontaneous polarization of pyroelectric BeO. Phys Rev Lett 64:1777–1780

Rabe KM, Ghosez Ph (2007) First-principle studies of ferroelectric oxides. In: Physics of ferroelectrics: a modern perspective. Topics in applied physics, vol 105. Springer, Berlin, pp 117–172

Resta R (1992) Theory of the electric polarization in crystals. Ferroelectrics 136:51–55

Resta R (1994) Macroscopic polarization in crystalline dielectrics: the geometric phase approach. Rev Mod Phys 66:899–915

Resta R (1998) Quantum mechanical position operator in extended systems. Phys Rev Lett 80:1800–1803

Resta R, Vanderbilt D (2007) Theory of polarization: a modern approach. In: Physics of ferroelectrics: a modern perspective. Topics in applied physics, vol 105. Springer, Berlin, pp 31–68

Resta R (2010) Electrical polarization and orbital magnetization: the modern theories. J Phys Condens Matter 22:123201

Resta R (2018) Drude weight and superconducting weight. J Phys Condens Matter 30:414001

Silvestrelli PL, Bernasconi M, Parrinello M (1997) Ab initio infrared spectrum of liquid water. Chem Phys Lett 277:478–482

Souza I, Iniguez J, Vanderbilt D (2002) First-principles approach to insulators in finite electric fields. Phys Rev Lett 89:117602

Spaldin NA (2012) A beginners guide to the modern theory of polarization. J Solid State Chem 195:2–10

Su WP, Schrieffer JR, Heeger AJ (1979) Solitons in polyacetylene. Phys Rev Lett 42:1698–1701

Thonhauser T (2011) Theory of orbital magnetization in solids. Int J Mod Phys B 25:1429–1458

Thonhauser T, Ceresoli D, Vanderbilt D, Resta R (2005) Orbital magnetization in periodic insulators. Phys Rev Lett 9:137205

Thonhauser T, Ceresoli D, Mostofi AA, Marzari N, Resta R, Vanderbilt D (2009) A converse approach to the calculation of NMR shielding tensors. J Chem Phys 131:101101

Umari P, Pasquarello A (2002) Ab initio molecular dynamics in a finite homogeneous electric field. Phys Rev Lett 89:157602

Vanderbilt D, King-Smith D (1993) Electric polarization as a bulk quantity and its relation to surface charge. Phys Rev B 48:4442–4455

Vanderbilt D, Resta R (2006) Quantum electrostatics of insulators – polarization, Wannier functions, and electric fields. In: Louie SG, Cohen, ML (eds) Conceptual foundations of materials: a standard model for ground- and excited-state properties. Elsevier, Amsterdam, pp 139–163

Wannier GH (1937) The structure of electronic excitation levels in insulating crystals. Phys Rev 52:191–197

Xiao D, Shi J, Niu Q (2005) Berry phase correction to electron density of states in solids. Phys Rev Lett 95:137204

Zak J (1989) Berry's phase for energy bands in solids. Phys Rev Lett 62:2747–2750

# Critical Phenomena in Glasses

# 9

Tommaso Rizzo

## Contents

**Abstract**

Supercooled liquids become increasingly sluggish upon cooling down to the glass temperature $T_g$ where they can no longer be studied in equilibrium on the laboratory scale and behave as off-equilibrium amorphous solids, i.e., glasses. Simple activated dynamics account for the behavior of so-called strong liquids, but deviations from Arrhenius behavior are observed in fragile ones and have defied explanation for decades. Technical advances in experiments have steadily unveiled more facets of the puzzling phenomenology of fragile liquids including notably two-step relaxation, stretched exponentials, superposition principles, and dynamical heterogeneities. Theoretical efforts have developed mainly around the idea that some sort of finite-temperature critical phenomenon is at play, the key role in the discussion being played by two different critical points. The first one is thought to occur above $T_g$, and therefore it is not really a phase transition

T. Rizzo (✉)
Institute for Complex Systems, Rome Unit, National Research Council of Italy, Rome, Italy

Physics Department, Sapienza University of Rome, Rome, Italy
e-mail: tommaso.rizzo@cnr.it

but rather a dynamical crossover. Numerical studies have shed much light on its nature, and nowadays it is largely believed to be the outcome of the smoothing of a sharp singularity spuriously predicted by mode-coupling theory. The existence of a dynamical crossover is largely accepted, and what is disputed is whether that is the end of the story.

Those who believe this is not the case typically put forward the classic hypothesis of a true thermodynamic phase transition to an amorphous glass state at some finite temperature below $T_g$. Originally suggested by elementary extrapolations of experimental data, this putative critical point is nowadays supposed to be a complex and fascinating object, notably the locus of a configurational entropy crisis accompanied by a divergent static correlation length. The quest to establish its existence, reinvigorated by the discovery of the glass/spin-glass analogy, is very much open but has produced nonetheless significant advances both at the theoretical and numerical level. Opponents of the thermodynamic transition scenario include notably those who advocate for dynamic facilitation, as realized in kinetically constrained models, to explain physics solely in terms of a dynamical crossover. Understanding dynamics between the crossover temperature and $T_g$ would help assess both the range of validity of a description in terms of the crossover and whether something qualitatively different must be invoked close to $T_g$ and below. Here the essential missing piece of information is the nature and spatial extent of the activated processes that should rule the dynamics: at the theoretical level, a consistent, beyond phenomenological, theory of these dynamical processes has still to be developed; at the experimental level, current techniques do not have enough spatial resolution; finally numerical simulations have been typically confined to higher temperatures due to hardware speed limitations but are beginning to access the crossover region and may provide some guidance in the coming years.

## 1    Introduction

In spite of decades of experimental, theoretical, and numerical studies, the glass problem is very much open: the community agrees only on a few statements while all the rest is strongly debated. In the following an introductory critical discussion will be given of some of the theoretical and numerical results inspired by the idea that some sort of critical phenomena determines the puzzling phenomenology of supercooled liquids. The reader should be aware that this covers only a part, although significant, of the literature of this field. In the spirit of the handbook, this contribution is not intended as a general and exhaustive review but rather as an, hopefully stimulating, introduction to some of the main open questions. The interested reader will find in the references a number of more technically detailed review papers and books (Gotze 2009; Cavagna 2009; Wolynes and Lubchenko 2012; Berthier et al. 2011; Berthier and Biroli 2011).

This line of research has evolved over the years around the existence of two critical temperatures: the crossover temperature and the Kauzmann temperature. The first critical temperature occurs upon supercooling the liquid when it still can be studied in equilibrium, and its existence is largely agreed upon and can be considered part of the phenomenology. The problem is that it does not mark a genuine phase transition but rather a dynamical crossover; thus it is difficult to make sharp and universal statements about it. In this context mode-coupling theory (MCT) captures the initial steps of the dynamical slowing down but fails spectacularly because it predicts a sharp transition instead of a crossover, and the main open problem is to extend it to the crossover region and below. From the point of view of material modeling and simulations, it is a particularly important problem, because the resolution of current experiments cannot provide guidance for its solution. On the other hand, microscopic details are fully under control in numerical simulation, but hardware constraints limit them to considerably higher temperatures than those reached by experiments. Nevertheless many believe that the crossover region is attainable with present or near-future technology and hope to finally understand this regime in the coming years.

It is fair to say that our shared understanding of the physics of glasses ends with MCT slightly before the glass crossover. Below it we only know that some kind of activated dynamic should set in. In this context a fascinating but strongly debated hypothesis is the occurrence, at very low temperatures, of a genuine thermodynamic transition from the liquid to an amorphous glass state. This line of research has inspired a variety of novel analytical and numerical techniques; the idea is that the information on slow dynamics can be obtained without performing molecular dynamic simulations but rather by studying the equilibrium configurations of appropriately constrained systems. From the analytical point of view, the advantage is that one can perform static computations that are typically easier than full-fledged dynamical ones. From the numerical point of view, the advantage is that equilibrium configurations can be obtained through more efficient algorithms than molecular dynamics. This is a field where numerical work has been going on in the last 15 years, and significant recent developments include algorithms to reach temperatures that even experiments cannot. Nevertheless evidences for a genuine phase transition are still not decisive; besides this information has not shed light into the fundamental question of why dynamics becomes so slow.

This work is organized as follows. In order to define the problem, the main dynamical features of realistic supercooled liquids as obtained from experiments will be summarized in Sect. 2. In Sect. 3 the MCT scenario will be presented, and it will introduce a discussion of the extent to which the phenomenology can be explained in terms of an avoided MCT-like transition. Section 4 is devoted to the existence of a genuine phase transition occurring in the regime that is not accessible by experiments and to the conceptual, technical, and numerical developments it has inspired. Conclusions will be drawn in the last section.

## 2    Phenomenology of Supercooled Liquids

The main properties of supercooled liquids, as observed in experiments, are (1) power law to exponential crossover, (2) two-step relaxation, (3) complex dynamics (stretched exponentials), and (4) dynamical heterogeneities (violation of the Stokes-Einstein relationship (SER)). These are dynamical properties that a consistent theory of glasses should necessarily explain. In addition there are the so-called thermodynamics/dynamics correlations whose significance in relationship to the experimental dynamical features is more disputed.

The power law to exponential crossover is the oldest property that has been observed and possibly the most important. Upon slowly cooling a liquid crystallizes at the melting temperature. If one is able to further lower the temperature avoiding crystallization, a supercooled liquid is obtained. Upon further cooling, the relaxation time or equivalently the viscosity of the supercooled liquid increases up to a value where the viscosity is equal to a (conventional) value of the order of the laboratory timescale. This point defines the glass transition temperature $T_g$. It should be clear that nothing special happens at $T_g$ from the point of view of the system; it is just the lowest temperature where the system can reach equilibrium in a timescale accept-able to humans. Obviously there is nothing strange in dynamic slowing down upon lowering the temperature; for instance, this occurs if dynamics proceeds through elementary events that require to overcome a constant free energy barrier $E$. In this case Arrhenius behavior is observed, with the logarithm of the relaxation time (or equivalently the inverse of the viscosity) increasing as the inverse of the temperature:

$$\tau_\alpha \propto e^{-\frac{E}{k_B T}} \tag{1}$$

Strong liquids exhibit by definition Arrhenius behavior upon supercooling, while fragile liquids exhibit a different behavior: when the logarithm of the viscosity is plotted vs. the inverse temperature in the so-called Angell's plot (1995), the data do not follow a straight line (see Fig. 1). In the early stages of supercooling, the viscosity of fragile liquids increases slowly and can be fitted by the tail of a power law diverging at some finite temperature. This temperature, however, marks a dynamical crossover: close to it the viscosity deviates from the power law and does not diverge; however it begins to increase in a more pronounced way that can only be fitted with functional forms depending exponentially on the temperature variations. Overall the data of fragile liquids exhibit a more or less pronounced bending with respect to the straight line Arrhenius behavior, and the slope at $T_g$ provides a quantitative definition of kinetic fragility (note that this terminology has nothing to do with the brittleness properties of the corresponding glass). Experimental data above $T_g$ have been fitted by many different functions, including some that, upon extrapolation, predict a divergence of the relaxation time at some finite temperature below $T_g$, notably the Vogel-Fulcher-Tammann (VFT) law:

$$\tau_\alpha = \tau_0 \exp\left[\frac{DT_0}{T - T_0}\right] \tag{2}$$
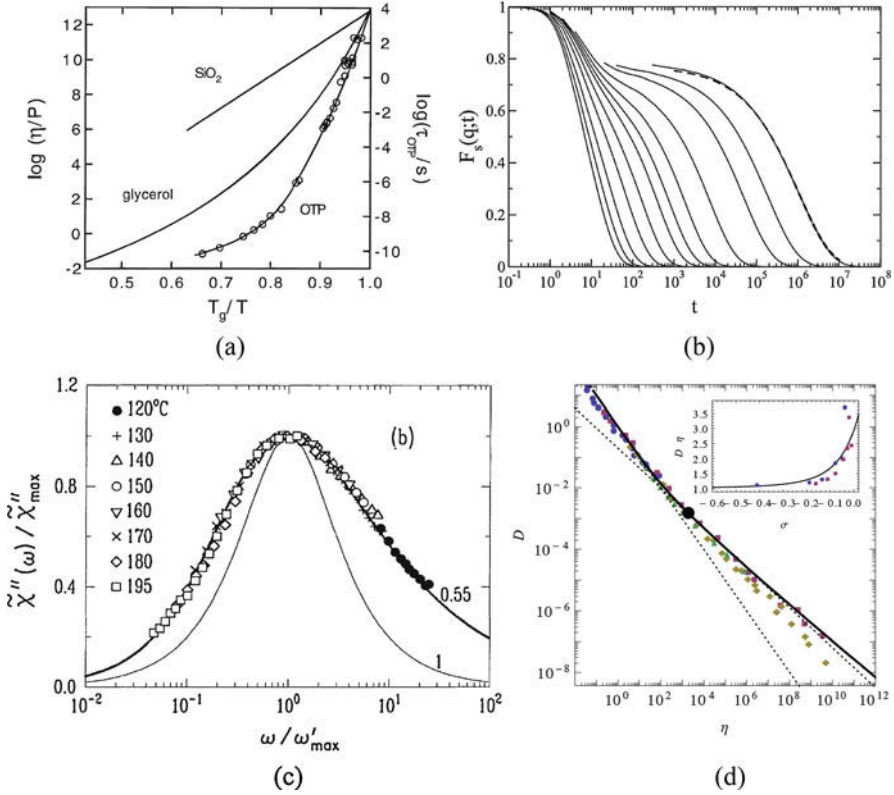
**Fig. 1** Phenomenology of supercooled liquids. (**a**) Strong liquids like silica ($SiO_2$) follow a straight line on the Angell's plot, while fragile liquids like glycerol and *ortho*-terphenyl (OTP) exhibit a bending (Ediger et al. 1996). (**b**) The self-intermediate scattering function $F_s(q; t)$ for $q = 6.1$ at various temperatures from a numerical simulation of a binary mixture, from Flenner and Szamel (2013). The dashed line is a stretched exponential. (**c**) The susceptibility spectrum of supercooled CKN (Li et al. 1992) is broader than that of a single exponential (solid line); correspondingly the relaxation can be only fitted by a stretched exponential. (**d**) Diffusion coefficient $D$ versus viscosity $\eta$ for experimental and numerical simulation data from Rizzo and Voigtmann (2015). Dashed lines indicate $D \propto \eta^{-1}$ (Stokes-Einstein relation) and a fit with $D \propto \eta^{-0.65}$ (fractional SE relation). A large circle marks the crossover temperature. Inset: $D \cdot \tau$ with simulation data

This leads to the two following essential questions: what causes the power law to exponential behavior in fragile liquids? Will the relaxation time eventually diverge at a finite temperature below $T_g$?

Models of supercooled liquids are often defined in terms of spherical particles interacting with pairwise potentials, and the typical dynamical observable is the intermediate scattering function:

$$F(q, t) \equiv \frac{1}{N} \langle \rho_q(t)\rho_{-q}(0) \rangle \tag{3}$$

where the square brackets mean thermal average, $\rho_q$ is the Fourier component of the density,

$$\rho_q \equiv \sum_{j=1}^{N} e^{i\,q\,r_j} \tag{4}$$

and $r_j$ is the position of particle $j$. The intermediate scattering function is equivalent, by means of Fourier transforms, to the dynamic structure factors and the van Hove function. At $t = 0$ the intermediate scattering function is equal to the static structure factor and relaxes to zero at large times. In experiments this quantity is accessible through a number of techniques (Richert 2012; Lunkenheimer et al. 2012). Observation of the two-step relaxation requires measurements on a huge range of timescales. The main relaxation, associated with the viscosity, is called the $\alpha$ process hence the name $\tau_\alpha$ for the relaxation time. However, there is an additional timescale $\tau_\beta$ over which $F(q, t)$ develops a plateau when plotted vs. $log(t)$ and the system appears to be essentially frozen. The $\beta$ scale is orders of magnitude smaller than $\tau_\alpha$, but it is considerably larger than the microscopic timescale. In Fig. 1b the phenomenon is displayed as measured in a numerical simulation for the relaxation of the self-intermediate scattering function (the Fourier transform of the single-particle displacement). The origin of the two-step relaxation is often attributed to the emergence of caging (Gotze 2009). The idea is that when dynamics is sufficiently slow, the environment of a given particle appears to be frozen, and therefore the particle itself is caged and will in turn cage its neighbors.

Caging is observed clearly in numerical simulations of hard-sphere systems (Kob 1999), but its origin is not completely understood. In these systems the temperature can be always reabsorbed into a rescaling of times, and the only nontrivial external parameter to be changed is the density. It seems natural that upon increasing the density at some point dynamics will slow down because of reduced volume effects. The problem with this interpretation is that caging is observed already at rather low density, and it appears abruptly in a rather narrow range of densities, meaning that if one inspects instantaneous configurations it is rather difficult to tell the difference between a configuration from a system that displays caging and one from a system that does not. This is an aspect of the so-called structure vs. dynamics problem. In other words the question is whether slow dynamics can be traced back to some structural property of the instantaneous equilibrium configurations. We may ask if there is some observable that can be measured on a given configuration that will correlate with the fact that relaxation from that configuration will exhibit caging or not. This is a fundamental question in the field, and the negative answer amounts to say that glassiness is a completely dynamical problem; this issue will be discussed again later on.

In simple Arrhenius dynamics relaxation decays exponentially in time. Relaxation in fragile liquids instead is complex, meaning that can only be typically

fitted with a stretched exponential form $\exp(-t^b)$. In experiments this result in the broadening of the susceptibility spectra that are related to the Fourier transform of the relaxation (see Fig. 1c). Much as the two-step relaxation, this is not at all trivial. As soon as the relaxation is controlled by independent microscopic events occurring in a uniform environment, the decay is necessarily exponential, while in order to have a stretched exponential behavior, one must either invoke spatial fluctuations in the environment or a moderately large correlation length implying that the relaxation events are not elementary or involve many particles. Which hypothesis is correct, if any of the two, is at present unknown. *Superposition principles* are an additional nontrivial feature of the relaxation. Strictly speaking they amount to say that the temperature dependence of the relaxation is only encoded in the timescale $\tau_\alpha$, meaning that one can write

$$F(q, t) = C_q(t/\tau_\alpha) \qquad (5)$$

for some function $C_q(t)$. This property is verified approximately by many supercooled liquids in some temperature range; see Gotze (2009) for a more detailed discussion of experimental data.

In the last 20 years dynamical heterogeneities (Ediger 2000; Berthier et al. 2011) have emerged as a fundamental property of fragile liquids. In short the idea is that approaching $T_g$ if one look at the liquid at the microscopic level there will be regions where dynamics is orders of magnitude faster than in the rest of the system. The above generic statement can, and often is, misinterpreted if one does not give it a more quantitative formulation. This is given by the so-called violation of the Stokes-Einstein relationship (SER) connecting the viscosity with the inverse of the diffusion constant of a given particle. The SER is verified at high temperatures, but violations are observed approaching $T_g$ and starting somewhere near the crossover temperature (see Fig. 1d). The SER can be justified assuming that each particle perform a kind of Brownian motion in a homogeneous environment. Conversely violations should occur in a situation in which, even if the SER is satisfied locally, the environment is not homogeneous and the local viscosity has strong fluctuations in such a way that the inverse of its average is different from the average of its inverse. Once again the precise mechanism leading to SER violations is not agreed on, but they are definitively considered a hallmark of complex dynamics. MCT helps to clarify that dynamical heterogeneities should not be confused with dynamical fluctuations; indeed MCT has diverging dynamical fluctuations but no SER violation.

In addition to the above dynamical features, a consistent part of the literature points toward the so-called thermodynamics/dynamics correlations connecting dynamical quantities like the viscosity with thermodynamical quantities (the entropy) measured by calorimetry. They are often invoked in the context of theories that advocate for the existence of a genuine phase transition below $T_g$, and they will be discussed in more details in Sect. 4.

# 3     The Crossover Temperature

## 3.1     Mode-Coupling Theory

Mode-coupling theory (Gotze 2009) starts from the exact microscopic equations for the dynamic structure factor of simple monoatomic liquid models and then makes a certain number of approximations. The outcome is a set of closed equations that can be solved numerically once the static structure factor is provided as input. Before discussing MCT predictions, let us stress that there is at present a very poor understanding of the approximations involved. The only justification comes a posteriori as the results display a significant agreement with the behavior of actual systems (again with some caveats), and the problem is complicated by the fact that some of the approximations involved cannot even be tested directly. Therefore, in spite of the fact that MCT starts from the exact microscopic dynamics, it is fair to say that it is essentially a black box whose internal operation is unknown. Efforts to understand it have been limited in the past, and more work should be devoted to this essential open problem. Instead research activity has essentially ignored this issue focusing on the empirical development of a sort of operating manual of the theory and applying the resulting procedure to a variety of models and systems again with impressing results.

There is a broad agreement that many supercooled fragile liquids can be modeled by simple liquids with pairwise interactions, including hard-sphere (HS) systems. However in two and three dimensions, monoatomic systems tend to crystallize easily and cannot be supercooled to very low temperature; therefore one must consider, at least, binary mixtures (Kob 1999). Mode-coupling equations can be extended to binary systems of hard or soft spheres, and the resulting equations can be solved to obtain the dynamic structure factor giving as input the static structure factor typically obtained from numerical simulation.

The theory predicts that at a given value of the control parameter (temperature or pressure), dynamical arrest occurs meaning that the infinite time limit of the dynamic structure factor does not decay to zero as it should in the liquid phase but tends to a constant momentum-dependent value, called the non-ergodicity parameter. As we said already it is known that this prediction is completely wrong, and in order to understand why, in spite of this dramatic failure, MCT is still considered highly valuable; we have to discuss its predictions upon approaching the transition temperature $T_{MCT}$.

First of all, MCT predicts that the decay of the dynamic structure factor proceeds in a two-step fashion as seen in experiments, with both scales $\tau_\beta$ and $\tau_\alpha$ diverging as power law of $\tau \equiv T_{MCT} - T$. More precisely at the critical temperature $T = T_{MCT}$, $F(q, t)$ approaches the nonzero long-time limit $F_c(q)$ in a power-law fashion with a nontrivial exponent $a$: $F(q, t) \approx F_c(q) + c/t^a$. The exponent $a$ controls the divergence of the $\beta$ scale through

$$\tau_\beta \propto |\tau|^{\frac{1}{2a}}. \tag{6}$$

For temperatures slightly larger than $T_{MCT}$, $F(q, t)$ reaches the plateau value $F_c(q)$ in a finite time that defines the scale $\tau_\beta$. For times of the order of $\tau_\beta$, it remains close to the plateau value up to deviations of order square root of $\tau \equiv T_{MCT} - T$ before eventually leaving the plateau on the scale $\tau_\alpha \propto |\tau|^{-\gamma}$. The decay on the $\alpha$ scale satisfies both the superposition principle and the stretched exponential form with exponent $\exp[-t^b]$, in striking agreement with experiments. Furthermore the exponent $\gamma$ is related to $a$ and $b$ through

$$\gamma = \frac{1}{2a} + \frac{1}{2b} \tag{7}$$

and the two exponents $a$ and $b$ can be both expressed as

$$\lambda = \frac{\Gamma^2(1-a)}{\Gamma(1-2a)} = \frac{\Gamma^2(1+b)}{\Gamma(1+2b)} \tag{8}$$

where $\lambda$ is the so-called parameter exponent for which MCT provides quantitative predictions.

Data from numerical simulations of a variety of models are consistent with this scenario in some range of temperatures (Kob 1999; Gotze 2009). In particular, it is often possible to estimate the critical temperature, the exponents, and the ergodicity-breaking parameter $F_c(q)$ and compare them with the quantitative predictions of MCT. One should bear in mind that these estimates from realistic systems data are intrinsically arbitrary because in principle MCT scalings are only well defined close to the dynamical arrest transition, but in practice the transition does not occur. In practice this ambiguity is reflected by the fact that the critical temperature depends on the observable and also on the nature and range of the fit. In spite of this intrinsic ambiguity, it turns out that the range of variations of the critical temperature is often sufficiently small for the whole procedure to make sense. This happens for most models and experimental systems, and indeed this *experimental* MCT critical temperature (or density) is almost always reported in any study of supercooled liquid. Furthermore the value of the ergodicity-breaking parameter can be obtained from the data, and it is often in excellent quantitative agreement with the predictions of MCT. The values of the critical exponents are more ambiguous but still are in good agreement with the predictions of MCT. In particular it seems that there is good consistency between the exponents $a$ and $b$ that control the $\beta$ regime with the exponent $\gamma$ that controls the $\alpha$ regime. This is a manifestation of the nontriviality of the double-step relaxation, in the sense that processes occurring on large but comparatively very different timescales are related. It is worth mentioning that MCT predictions appear to be somewhat meaningful even below the critical temperature $T_{MCT}$ where the theory predicts that $F(q, t)$ have a nonzero long-time limit increasing as $\sqrt{\tau}$ for $T$ smaller than $T_{MCT}$. It turns out that in many experimental systems and numerical simulations a square root increase of the plateau value can be actually observed notwithstanding the fact that $F(q, t)$ will eventually leave the plateau at variance with MCT predictions.

Overall the comparison between numerical simulations and theory suggests a scenario in which ideal MCT predictions are accurate except very close to the critical temperature where the theory needs some substantial modification. Unfortunately this scenario is too optimistic in the sense that the effective MCT temperature, the one for which MCT fits do describe actual data, is definitively different from the value of the critical temperature obtained from the solutions of the MCT equations itself (Kob 1999). This discrepancy is model-dependent but seems rather universal; typically it is more pronounced in soft sphere models than in HS, but its origin is unclear. As a result, it is often said that the MCT critical temperature is wrong, but it should be clear that there are two essentially different aspects of it being wrong. In a weak sense it is wrong because its value does not compare well with the value used to fit numerical data. The theory overestimates this value, and in practice in order to obtain reasonable fits, one must plug into the MCT equations the structure factors from higher temperatures. In a strong, more fundamental, sense MCT is wrong because there is no true critical temperature in experiments and numerical simulations.

The above discussion on the critical temperature allows to introduce the quantitative vs. qualitative MCT scenario. There are some features of MCT that are more universal and are found also outside the original domain of the theory, i.e., supercooled simple liquid models in physical dimension. The most striking instance is the case of spin glasses displaying one step of replica symmetry breaking (1RSB) in Parisi's scheme (Mezard et. al. 1988). These systems are utterly different from a liquid at the microscopic level; notably they are defined on a lattice, and they have quenched disordered interactions, and yet, as discovered by Kirkpatrick and Thirumalai (1987), their dynamics shares the very same phenomenology of MCT, namely, dynamic arrest characterized by two-step relaxation with power-law divergent timescales. Furthermore in the $\beta$ regime, where the correlation stays near the plateau, dynamics obeys the very same critical equation of MCT, i.e., a universal quadratic dynamical equation that does not depend on the microscopic details of the model. Thus one should always bear in mind that there are qualitative, much more general, MCT predictions and quantitative MCT predictions that are only limited to supercooled liquid models. In MCT literature this distinction is often not very clear.

An important instance of the qualitative/quantitative difference is the case of simple supercooled liquids in large dimension. This problem has been investigated extensively over the years under the expectation that in infinite dimension the crossover should become a true transition and the problem should become solvable. These studies culminated recently with the exact and complete solution (Charbonneau et al. 2016) that confirmed that the qualitative MCT scenario is correct (e.g., the relationship between the exponents holds), while the quantitative values of the critical temperature and of the exponents are wrong.

Another striking instance of the universality of the qualitative MCT scenario is provided by cooperative kinetically constrained models (KCM) (Ritort and Sollich 2003; Chandler and Garrahan 2010) that will be briefly discussed in the next

section. Remarkably the Fredrickson-Andersen model, when studied on the (mean-field) Bethe lattice, displays a dynamical arrest transition in agreement with the qualitative MCT scenario (Berthier and Biroli 2011). Note that this statement is only based on numerical observations, and justifying it analytically is a fascinating open problem.

We are now in position to discuss the failures and successes of MCT in connection with the phenomenology of real supercooled liquids presented in the previous section. The two-step relaxation scenario is a significant success together with the presence of stretched exponentials in the relaxation. The main failure is the prediction of a true phase transition which in reality is a crossover; however MCT is at least consistent with the initial power-law behavior. Another significant failure of MCT is that even close to the critical temperature, it does not display the SER violations associated with dynamical heterogeneities. By means of the microscopic MCT equations, one can indeed obtain predictions for the diffusivity. Remarkably the lack of SER violation is confirmed by the exact solution in infinite dimension (Charbonneau et al. 2016). The fact that SER holds in MCT does not mean that dynamical fluctuations are finite within the theory. Actually one can argue that close to $T_{MCT}$ the theory predicts both diverging dynamical fluctuations and diverging dynamical correlation length. This can be shown by means of an extension called inhomogeneous MCT (Berthier and Biroli 2011) although care must be taken in extracting the critical exponents as simple scaling arguments lead to incorrect results (Rizzo 2014). Given that the dynamical arrest transition does not occur, these divergences are also unrealistic and are not often emphasized in the MCT literature.

A few comments on the spurious divergence of the dynamical correlation length are in order. In the early days of MCT, it was not clear what was the origin of dynamical arrest, and some authors believed that no diverging correlation length was involved (see, e.g., Ediger et al. 1996). One can argue on physical ground that for ergodic statistical mechanics systems, a finite-temperature dynamical transition must be necessarily associated with a diverging correlation length (these arguments can be made rigorous in the case of spin-glass models (Berthier and Biroli 2011)). This observation may appear not so important in the case of the MCT crossover because in the end no dynamical arrest transition is present, but it is important when discussing the possibility of a true transition occurring at some finite temperature below $T_g$ as we will see in the following section. Another point worth mentioning is that while diverging fluctuations must be necessarily accompanied by a diverging correlation length, a simple increase of fluctuations does not imply an increasing correlation length. This is important to assess the relevance of experimental measurements of nonlinear susceptibilities. The increase of fluctuations reported in these experiments is sometimes interpreted as an indication of an increase of the correlation length (Albert et al. 2016), overlooking the fact that the connection between the two quantities requires additional assumptions. Actually a recent theory of the crossover provides a counterexample in which increasing dynamical fluctuations are accompanied by a decreasing correlation length (Rizzo 2014).

## 3.2    Toward a Theory of the Crossover

The understanding of supercooled liquids ends slightly above the crossover temperature and everything occurring below is disputed. While MCT is definitively wrong at low temperatures, its success at moderate supercooling suggests that one should amend it in order to describe the crossover region and then move to even lower temperatures. Efforts in this direction have been going on for a long time, but none is considered fully satisfactory and will not be discussed in detail. People tried often to modify the theory at the microscopic level. In general it turns out to be rather easy to remove the sharp transition, but this typically leads to a simple exponential relaxation thus spoiling the stretched exponential which is one of the great successes of MCT, not to mention that these modifications are always ad hoc and phenomenological. Note that MCT also starts from the microscopic first-principle description and then makes some uncontrolled approximations, however there is no evident logical connection between the approximations made and the final results, and in this sense MCT approximations are not considered ad hoc.

A different possibility is that in order to describe the physics at the crossover and go beyond MCT, one should use instead a mesoscopic effective theory that is accurate on a coarse-grained scale over which the microscopic details of the model are unimportant. This is the way to proceed in the case of a genuine second-order phase transition because they are characterized by a divergent correlation length. Applications of these ideas in the context of the crossover appear counterintuitive, since in the end the correct theory should tell us that there is no divergence of the correlation length. Nevertheless one should note that in order for a mesoscopic effective theory description to be valid, one does not need a diverging correlation length; it is enough for it to be large compared to the microscopic scale. On the other hand it is largely agreed that dynamics in the crossover region is cooperative meaning that the dynamical correlation length, while not diverging, is indeed large (Berthier and Biroli 2011; Harrowell 2011).

The mesoscopic theory can be shown to be equivalent to a solvable model of dynamical stochastic equations. Quite interesting the resulting model displays all the hallmark of supercooled liquid, namely, the power law to exponential crossover, two-step relaxation, stretched exponential, and dynamical heterogeneities in the form of SER violations (Rizzo 2014). Most importantly it is not a phenomenological theory because these features are not the outcome of ad hoc assumptions. On the other hand, by definition, there is no way to tell from the mesoscopic theory itself if it is appropriate to describe a specific system. For that one should start from a first-principle microscopic description and apply coarse-graining in a rigorous way. This implies that in principle this description could be accurate for one system but not for another system. It is also important to note that the validity of a mesoscopic description for a given supercooled liquid model could also be assessed explicitly by means of a numerical simulation. This requires to accurately measure dynamical fluctuations and correlation lengths to check if the order parameter after coarse-graining is sufficiently smooth to be described by an effective theory.

The technology to measure dynamical correlations and fluctuations was developed some 20 years ago motivated by the experimental observations of dynamical heterogeneities (Berthier and Biroli 2011), and it is known that extracting these quantities from numerical simulations is quite demanding.

## 4 Dynamics vs. Thermodynamics: The Kauzmann Temperature

The idea of a genuine phase transition occurring below originated with Kauzmann's analysis of calorimetric experimental data (Kauzmann 1948). He noticed that while at the melting temperature the entropy of the crystal is lower than that of the liquid, it decreases less upon cooling in such a way that, by extrapolation, they should become equal at a temperature $T_K < T_g$. In order to avoid the paradox of an amorphous state having lower entropy of the crystal, he considered various possibilities including that the liquid had a genuine phase transition to a thermodynamically stable glass state at $T_K$. Nowadays Kauzmann's observation is not considered a paradox anymore: in the late 1950s computer simulations showed, to the surprise of many, that an isolated liquid of hard spheres does crystallize, and this occurs precisely because the crystal state has a higher entropy than the disordered state! This is understood in terms of a balance between short-range and long-range entropies: the crystal has a smaller long-range contribution to the entropy but has a larger local contribution because the particles have locally more room than in the amorphous state (Ackerson 1993). As we said already, crystallization is actually a major problem in numerical simulations of supercooled liquids, and it is typically avoided using binary mixtures or more polydisperse systems.

In spite of the fact that Kauzmann 's observation is not really a paradox, the debate over the existence of a genuine phase transition is still going on to this day. Interestingly Kauzmann himself was against this hypothesis and believed that the paradox would be solved by some dynamical mechanism. The main problem being that it is not possible to study the supercooled liquid below $T_g$. Another important aspect of the discussion is the observation of thermodynamics/dynamics correlations (Adam and Gibbs 1965). Adam and Gibbs pointed out that for many experimental systems VFT fits of viscosity data yield a critical temperature $T_0$ compatible with the $T_K$ extrapolated from calorimetric data. In particular if one plots the logarithm of the viscosity (a dynamical quantity) as a function of the inverse of the excess entropy , i.e., the difference between the entropy of the supercooled liquid and that of the crystal (a thermodynamic quantity), they appear to be correlated.

To complete the picture, Goldstein (1969) started with the observation that approaching $T_g$ relaxation from a given initial configuration proceeds in a two-step fashion; first each particle explores its cage and then escapes from it. Then he argued that the motion inside a cage defines a sort of metastable state and the final relaxation can be seen as a jump from one metastable to another. If one further argues that the entropy of the metastable state corresponds to a kind of vibrational

contribution (essentially analogous to the crystal entropy), it follows that the excess entropy is a measure of the number of metastable states, called the configurational entropy. The decrease of the configurational entropy thus explains the slowing down of dynamics because the system in order to relax has less and less metastable states to escape to.

It is important to realize that while the excess entropy is a well-defined object that can be measured experimentally, the configurational entropy is not. Entropy is a static equilibrium concept and thus cannot be applied to a metastable state that can only be observed on a finite time window. More practically this is reflected by the fact that there is no unique definition of it. On the other hand, Kirkpatrick, Thirumalai, and Wolynes (KTW) pointed out that the previously mentioned mean-field spin-glass models are indeed characterized by the presence of many metastable states below a temperature $T_{MCT}$ and the corresponding configurational entropy vanishes at some lower temperature $T_K$ (Kirkpatrick and Thirumalai 1987). At the mean-field level there is simply no dynamics beyond $T_{MCT}$, and they put forward the idea that in physical dimension the mean-field picture should be modified invoking nucleation arguments as in the classic Becker-Doring theory.

Classical nucleation theory describes the decay of a metastable phase to the stable phase of lower free energy in terms of the expansion of droplets of the stable phase. The surface tension tends to shrink the droplet, while the bulk free energy difference tends to expand it. The two forces depend on the radius of the droplet and balance at some critical radius $r_c$: droplets larger than $r_c$ will expand, while smaller droplets will shrink and disappear. The time to nucleate a critical droplet naturally increases with the size of the critical radius which in turn increases and diverges at the coexistence point where the free energy difference is zero and both phases are stable; correspondingly the lifetime of the metastable state diverges exponentially. The idea of KTW is to use nucleation arguments to describe the dynamics of supercooled liquids using the configurational entropy in place of the free energy difference. This offers an explanation of the thermodynamics/dynamics correlations and of the exponential VFT-like divergence of the relaxation time. Furthermore the nucleation argument implies that there is a diverging correlation length, which was already advocated by Adam and Gibbs that suggested that dynamics is driven cooperatively rearranging regions (CRR) of increasing size upon approaching $T_g$.

The existence of an actual phase transition at $T_K$ is an appealing topic because it is a sharp statement, either true or false. On the other hand the nontrivial phenomenology we need to explain is observed *above* $T_g$, and thus a consistent part of the community does not consider it particularly relevant. At any rate, even if a genuine glass transition occurred below $T_g$, one should still have to prove that the physics observed above $T_g$ is a consequence of its presence. Instead many believe that the physics of supercooled liquid can be explained solely in terms of a crossover temperature that was discussed in the previous section. Note that in the context of mean-field theories, the existence of a $T_K$ requires logically the presence of a crossover at a higher temperature while the opposite is not true: models exist for which there is an MCT singularity, but the configurational entropy does not vanish at any finite temperature.

The opposite view that dynamics may be completely unrelated to thermodynamics is supported by the so-called Kinetically constrained models (Ritort and Sollich 2003; Chandler and Garrahan 2010). They are defined in terms of binary variables (that can be viewed as spins) on a lattice. The variables do not interact and thus the thermodynamics is trivial at all temperatures. Dynamics instead must obey some constraint, for instance, in the well-known Fredrickson-Andersen model, each spin must have at least $k$ neighbors in the up state in order to flip. It turns out that dynamically these models display many of the phenomenological features of supercooled liquids, notably a power law to exponential crossover and two-step relaxation. Since no thermodynamics transition may occur in these models, their physics is typically interpreted solely in terms of the existence of a crossover temperature. Furthermore on the Bethe lattice they seem to display a sharp MCT-like transition, and the aforementioned dynamical features can clearly be associated with this transition being avoided on lattices in three and two dimensions (Berthier et al. 2011). The main open problem is that these models lack a clear microscopic connection with the original liquid models, and it is not exactly clear how the constraints on the dynamics should emerge from the (unconstrained) microscopic Hamiltonian dynamics. In practice this is also reflected by the impossibility of obtaining some sort of quantitative prediction. Another problem is that it seems difficult to rationalize the Adam-Gibbs phenomenological correlations between dynamics and excess entropy using these models, although, as we said before, one may question their overall relevance.

An additional interesting open problem is inherently associated with facilitation dynamics. At low temperatures one expects that in a supercooled liquid there are large regions that are essentially frozen for a large amount of times and eventually relax. One possible mechanism inducing the relaxation is the expansion/motion of a mobile region from outside the blocked region into it, but one can also think of relaxation led by thermally activated rare events. However this second mechanism is impossible in KCM where relaxation can only propagate from the borders of the blocked region. Thus knowing what is the mechanism at work in actual supercooled liquids could help discriminate between competing theories, but at present this is unclear.

From a practical point of view, the idea of a genuine thermodynamics transition is appealing because one can study it in a purely static framework. In the context of numerical simulations, one is no longer bound to use a physical algorithm to thermalize the system, and more efficient algorithms can be devised. For instance, particle-swap algorithms have been used in the last 15 years and recently have produced spectacular results for specially designed polydisperse hard-sphere mixtures that were thermalized up to densities corresponding to the laboratory timescale and even below (Berthier et al. 2017).

Underlying any static study in the context of supercooled liquids is the key question of whether nontrivial information can be extracted solely from the instantaneous equilibrium configurations. At first one would say that the answer is no, after all one of the striking features of supercooled liquid models of hard or soft spheres is precisely the fact that if one looks at the instantaneous configurations

at various temperatures (or densities) approaching the crossover, one cannot detect any significant change that could justify the dramatic increase of the relaxation time. Nevertheless, guided by the supercooled liquid/spin-glass analogy advocated by KTW, various authors have suggested that nontrivial information can be obtained solely from the equilibrium configurations by applying appropriate procedures (Wolynes and Lubchenko 2012).

The general idea is that one has to use an equilibrium configuration of the original system (the reference configuration) to define a new system and then show that nontrivial information can indeed be red off the equilibrium configurations of the new system. The first procedure that was introduced is the so-called Franz-Parisi potential that counts the equilibrium configurations that are at some fixed distance from the reference configuration. At the mean-field level, the potential has a minimum as a function of the distance at the maximal distance but, upon lowering the temperature, develops a secondary minimum at $T_{MCT}$. The difference in height between the primary and the secondary minimum is the configurational entropy that will eventually vanish at a lower temperature . This mean-field picture must be modified in finite dimension by applying a Maxwell construction as in ordinary first-order phase transition: one expects that the potential first develops a constant-slope segment close to the crossover temperature with the slope eventually going to zero at $T_K$.

Another interesting procedure is to measure the so-called point-to-set length (Berthier and Biroli 2011). In this case the particle positions are frozen in the reference configuration except inside a spherical cavity that it is then thermalized again. One expects that when the cavity is very large the bulk should be insensitive to the constraint at the boundaries, while if the cavity is sufficiently small, the center of the cavity should be stuck in a metastable state (stabilized by the frozen boundary). The point-to-set length is defined as the length that separates these two regimes, and mean-field theory predicts that it should diverge at $T_K$. This length should provide a way to actually measure the size of CCRs thus rephrasing the KTW idea of a diverging correlation length in a way that is amenable to be tested in simulations.

Finally we can mention random pinning that amounts to freeze a finite fraction of the particles of the reference configuration and equilibrate again the remaining free particles (Berthier and Biroli 2011). In this case the existence of genuine phase transition can be linked to the presence of a line of first-order transitions in the temperature-concentration plane. Different geometries, e.g., the freezing/pinning of an infinite wall of the system, have also been studied. Note that in both the pinning and cavity procedure, it can be shown that the configuration in which the free particles have the same positions they had in the reference configuration is an equilibrium configuration for the constrained system, which is a considerable advantage because one does not need to equilibrate the system again.

In the last 20 years the above procedures have been applied to various systems, and the most spectacular results are those obtained recently by the swap algorithms (Berthier et al. 2017). In these studies both the Franz-Parisi potential and the

point-to-set length were measured in equilibrium for values of the density even higher than the glass density $\rho_g$. Their behavior was found to be compatible (through extrapolation) with the presence of an actual phase transition at some $\rho_K$. Unfortunately, as it will be discussed in the following, while being a tremendous improvement with respect to earlier studies, it seems that these results are not enough to assess convincingly the validity of the scenario.

One problem is that the degree of extrapolation that one has to make on the actual data is still significant, and if one thing can be learned from the extensive literature on glasses is that extrapolations are always debated. Another problem concerns the point-to-set length itself and emerged from the early numerical studies of this quantity. Measurements of this quantity indicate a rather small value that increases slowly; thus, while a divergence at lower temperature cannot be ruled out, it is not particularly remarkable. Given that the growth and eventual divergence of this quantity should be the driving mechanism behind the growth of the relaxation time, we have a problem because this length is small while the relaxation time is increasing exponentially. Therefore it is further postulated that the relationship between the static length and the relaxation time is exponential due to some kind of activation mechanism. To this day however, a solid theoretical foundation of these statements is lacking.

The connection with dynamic raises an additional problem. As we discussed in the previous section, the dynamical correlation length should diverge at $T_{MCT}$. In the region where it can be measured, it exhibits a significant growth, while obviously being nondivergent. However in the region where both can be measured, the dynamical correlation length is definitively larger than the point-to-set length and appears to be growing more strongly, such that extrapolation suggests that their difference would be even more pronounced at lower temperatures. In other words the static length does not seem to be relevant for dynamics in the region where dynamics is already nontrivial.

In order to complete the discussion on the relevance of static methods, one must add that unphysical algorithms may allow to assess the validity of supercooled liquid models. As we said already the problem of crystallization becomes less severe increasing the dimension, but in three and two dimensions one is necessarily forced to consider at least binary mixtures. Recent numerical studies suggest that even if they do not crystallize at the temperatures reached by state-of-the-art molecular dynamics simulations they will be at lower temperatures accessible at present only by unphysical algorithms. Furthermore while up to now the knowledge of low-temperature equilibrium configurations has not shed light into the central problem of why dynamics slows down, they could be used to study the emergence of various anomalies in the glassy state observed in experiments. Besides understanding why swap algorithms are successful may shed light into why physical dynamics is instead so slow.

Some comments are important on the analogy between spin glasses and super-cooled liquids suggested almost 30 years ago by KTW. On general ground systems with a completely different microscopic structure can be expected to share some

common behavior if they have a sufficiently large correlation length, such that the relevant physics occurs on a scale insensitive to microscopic details. In this light the fact that static length measured is typically small is troublesome. Furthermore many problems concerning these spin-glass models are still open irrespective of their eventual connection with supercooled liquids. First of all there is no numerical evidence that any of these spin models actually displays an entropy crisis transition with a discontinuous order parameter in finite dimension: the fate of the MF transition in finite dimensional system has still to be firmly established. Various failures to identify such a model have been rationalized by noticing that short-length fluctuations, already present on finite-connectivity mean-field lattices, are responsible for the absence of the transition rather than long-length fluctuations that would be a more general and harmful mechanism. Thus the quest for a good candidate is still a fascinating open problem.

Besides the numerics, an additional problem is that there is no precise and well-established analytical treatment of finite dimensional effects. In particular the connection between statics and activated dynamics is not well established even at the mean-field level, i.e., that of nucleation theory for metastability and phase coexistence. Furthermore, in order to agree with experimental data, the phenomenological expressions require the use of non-mean-field exponents that nobody knows at present how to compute. Note that these are all well-defined problems that should be addressed and solved irrespectively of their relevance to supercooled liquids.

We have seen that the static approaches offer a considerable advantage at the algorithmic level, but there are also major advantages at the analytic level, because dynamics is typically more difficult than statics. The spin-glass/structural glass analogy suggests that one has to use the replica method (Mezard and Parisi 2012) that indeed provides a way to compute static objects like the Franz-Parisi potential and the point-to-set length. In finite dimension the replica method predicts a MCT-like transition that should be considered a spurious mean-field modification of what is in reality a crossover. This follows from the fact that, as in MCT, one must resort to some approximation scheme. As we said before, MCT provides nevertheless good quantitative estimates for quantities like the ergodicity-breaking parameter and the (pseudo)critical exponents. On the contrary current approximation schemes in the replica method studies yield predictions of considerable lower quality. On the other hand, as we said in the previous section, the replica method has been used successfully to study supercooled liquids in infinite dimension. This is an instance, albeit special, of a system where the connection between dynamics and thermodynamics can be demonstrated starting from first-principle microscopic methods, to be contrasted with KCM models, where, at present, there is no explicit microscopic derivation of the assumption that dynamics is facilitated. Still, the essential problem remains to describe how the ideal MCT transition becomes a crossover in finite dimension, and it is not even clear if the mechanism at work in large but finite dimension is the same in dimension two and three.

# 5 Conclusions

The body of work presented in the previous sections may be summarized as follows: there is an accepted transition which is not really a genuine critical point, and then there is another transition that is really critical but whose existence is highly debated.

State-of-the-art molecular dynamics simulations can equilibrate a few super-cooled liquid models down to temperatures close and slightly below the crossover temperature. While there is no precise way to substantiate these claims as we are dealing with a crossover and not a sharp critical point, there is, nevertheless, a widespread belief in the community that the next generation of numerical studies will be able to probe the crossover region and beyond for a variety or models. This makes all the more urgent to go beyond MCT and develop a comprehensive theory of the crossover with qualitative and quantitative predictive power.

Additional motivation to tackle this problem comes from other critical phenomena that have not been discussed here. In particular in the context of MCT, higher-order glass-glass transitions have been predicted and also detected numerically to some extent (Gotze 2009). Actually, the ability to predict these singularities prior to observation has convinced many of the value of MCT and of the fact that one should try to correct its shortcomings instead of throwing it away altogether. Much as the MCT transition, these higher-order singularities are expected to become smooth crossovers in realistic systems, and it seems reasonable that their understanding would benefit from any development in the former. Similarly, the study of liquids in high dimension (Charbonneau et al. 2016) has suggested the existence of even more exotic critical points, notably full replica-symmetry-breaking critical points as observed in mean-field spin glasses. At present the relevance of these transitions in realistic models is unclear. Be as it may, they are intrinsically mean field in nature, and their existence in realistic models poses a number of conceptual problems. Actually they may be observed only after the MCT crossover has occurred, and it is likely their fate in finite dimension could be fully understood only once the crossover problem has been solved.

As for the debate on the existence of a genuine glass transition below $T_g$, the bottom line is that the static objects that should manifest critical behavior, e.g., static length scales, are compatible with a singularity, but their divergence is too mild and not particularly remarkable in the range where they can be measured in current numerical simulations. On the other hand their connection with the truly remarkable dynamical features is still speculative and should be established more rigorously. Nevertheless this is a field in which progress has been made steadily in the last decades and more is expected to come.

If we had to single out one fundamental open question from the previous discussion, that would be the nature of the activated process below the crossover temperature and above $T_g$. Measurements of increasing dynamical correlation lengths unambiguously show that kinetic fragility, i.e., the power-law-to-exponential crossover is induced by the dynamics becoming cooperative. The problem is what happens at lower temperatures: do correlation lengths continue to increase? Can dynamics

be described solely in terms of the crossover, or can we identify some feature that necessarily requires some other mechanism approaching $T_g$? Currently, experiments do not have enough spatial resolution to shed light on these questions while the timescales involved are too large for numerical simulations, but both problems could be overcome in the future through some breakthrough.

# References

Ackerson BJ (1993) When order is disordered. Nature 365:11–12

Adam G, Gibbs JH (1965) On the temperature dependence of cooperative relaxation properties in glass-forming liquids. J Chem Phys 43(1):139–146

Albert S, Bauer T, Michl M, Biroli G, Bouchaud J-P, Loidl A, Lunkenheimer P, Tourbot R, Wiertel-Gasquet C, Ladieu F (2016) Fifth-order susceptibility unveils growth of thermodynamic amorphous order in glass-formers. Science 352:1308

Angell CA (1995) Formation of glasses from liquids and biopolymers. Science 267:1924

Berthier L, Biroli G (2011) Theoretical perspective on the glass transition and amorphous materials. Rev Mod Phys 83:587

Berthier L, Biroli G, Bouchaud J-P, Cipelletti L, van Saarloos W (eds) (2011) Dynamical heterogeneities in glasses, colloids, and granular media. Oxford University Press, Oxford

Berthier L, Charbonneau P, Coslovich D, Ninarello A, Ozawa M, Yaida S (2017) Configurational entropy measurements in extremely supercooled liquids that break the glass ceiling. PNAS 114:11356–11361

Cavagna A (2009) Supercooled liquids for pedestrians. Phys Rep 476(4–6):51–124

Chandler D, Garrahan JP (2010) Dynamics on the way to forming glass: bubbles in space-time. Annu Rev Phys Chem 61:191

Charbonneau P, Kurchan J, Parisi G, Urbani P, Zamponi F (2016) Glass and Jamming transitions: from exact results to finite-dimensional descriptions. Annu Rev Condens Matter Phys 8: 265–288

Ediger MD (2000) Spatially heterogeneous dynamics in supercooled liquids. Annu Rev Phys Chem 51:99

Ediger MD, Angell CA, Nagel SR (1996) Supercooled liquids and glasses. J Phys Chem 100(31):13200–13212

Flenner E, Szamel G (2013) Dynamic heterogeneities above and below the mode-coupling temperature: evidence of a dynamic crossover J Chem Phys 138:12A523

Goldstein M (1969) Viscous liquids and the glass transition: a potential energy barrier picture. J Chem Phys 51:3728–3739

Götze W (2009) Complex dynamics of glass-forming liquids: a mode-coupling theory. Oxford University Press, Oxford

Harrowell P (2011) The length scales of dynamic heterogeneity: results from molecular dynamics simulations. In: Berthier L, Biroli G, Bouchaud J-P, Cipelletti L, van Saarloos W (eds) (2011) Dynamical heterogeneities in glasses, colloids, and granular media. Oxford University Press, Oxford

Kauzmann AW (1948) The nature of the glassy state and the behavior of liquids at low temperatures. Chem Rev 43:219

Kirkpatrick TR, Thirumalai D (1987) Dynamics of the structural glass transition and the -Spin—Interaction Spin-Glass Model. Phys Rev Lett 58:2091

Kirkpatrick TR, Wolynes PG (1987) Connections between some kinetic and equilibrium theories of the glass transition. Phys Rev A 35:3072

Kob W (1999) Computer simulations of supercooled liquids and glasses. J Phys Condens Matter 11:R85

Li G, Du MN, Chen XC, Cummins HZ, Tao NJ (1992) Testing mode-coupling predictions for $\alpha$ and $\beta$ relaxation in $Ca_{0.4}K_{0.6}(NO_3)_{1.4}$ near the liquid-glass transition by light scattering. Phys Rev A 45(6):3867

Lunkenheimer P, Kohler M, Kastner S, Loidl A (2012) Dielectric spectroscopy of glassy dynamics. In: Wolynes PG, Lubchenko V (eds) Structural glasses and supercooled liquids: theory, experiment, and applications. Wiley, Hoboken

Mezard M, Parisi G (2012) Glasses and replicas. In: Wolynes PG, Lubchenko V (eds) (2012) Structural glasses and supercooled liquids: theory, experiment, and applications. Wiley, Hoboken

Mezard M, Parisi G, Virasoro M (1988) Spin glass theory and beyond. World Scientific, Singapore

Richert R (2012) Supercooled liquid dynamics: advances and challenges. In: Wolynes PG, Lubchenko V (eds) Structural glasses and supercooled liquids: theory, experiment, and applications. Wiley, Hoboken

Ritort F, Sollich P (2003) Glassy dynamics of kinetically constrained models. Adv Phys 52: 219–342

Rizzo T (2014) Long-wavelength fluctuations lead to a model of the glass crossover. EPL 106:56003

Rizzo T, Voigtmann T (2015) Qualitative features at the glass crossover. EPL 111:56008

Rizzo T (2016) Dynamical Landau theory of the glass crossover. Phys Rev B 94:014202

Wolynes PG, Lubchenko V (eds) (2012) Structural glasses and supercooled liquids: theory, experiment, and applications. Wiley, Hoboken

# Part III
# Electronic Structure of Materials by Ab Initio Methods

# Electronic Structure of Materials by Ab Initio Methods: Overview

# 10

## Angel Rubio

## Contents

**Abstract**

The next set of 12 chapters provides an overview of the new advances since the first edition of the *Handbook of Materials Modeling* in 2005 regarding the description of the ground-state and excited-state electronic structure of complex many-body systems by ab initio electronic structure methods. In this section we present contributions aiming to providing an up-to-date description and illustration of the main theoretical methods used by the electronic structure community for the study of problems of actual materials, of prediction of properties, and for the design of novel materials.

## 1     Introduction

Back in September 2015, I was contacted by Wanda and Sidney, who proposed me the idea of joining the editorial board of the second edition of the *Handbook*

A. Rubio (✉)

Theory Department, Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

Center for Computational Quantum Physics (CCQ), The Flatiron Institute, New York, NY, USA
e-mail: angel.rubio@mpsd.mpg.de

*of Materials Modeling* as well as being the editor in charge of one section on electronic structure and ab initio methods, which should work as a fundamental point of reference for the following chapters within the volume. I thought this was an excellent idea. After a lot of hard and excellent work by the contributors to this section, I am delighted to welcome you to the introduction of the section on "Electronic structure of materials by ab initio methods" in the *Handbook of Materials Modeling (2nd edition, Vol. 1)*.

A large part of research in condensed matter science is related to the characterization of the electronic, structural, and bonding properties of interacting many-electron systems. An accurate description of the electronic structure and its response to external probes is essential to the understanding of the behavior of systems ranging from atoms, molecules, and nanostructures to complex materials. Solving for the electronic structure of an interacting electron system (in terms of the many-particle Schrodinger equation) has an intrinsically high complexity: while the problem is completely defined in terms of the total number of particles N and the external potential V(r), its solution is a task of exponentially increasing complexity. Fortunately, in the study of either ground- or excited-state properties of interacting many-body systems, we seldom need the full solution to the Schrodinger equation. When one is interested in structural properties, the ground-state total energy is sufficient. In other cases, we want to study how the system responds to some external probe; thus, the knowledge of a few excited-state properties is needed.

In the last two decades, we have seen a continuous but impressive improvement in the accuracy of the approximate methods to deal with many-body electronic systems together with a tremendous increase of computing power as well as improved numeral algorithms. In this realm, ab initio computational sciences have become a fundamental tool to understand, discover, and predict materials properties, as well as to guide experimental efforts along those lines. Therefore, in this set of chapters, we aim to lay out the fundamental concepts providing the basis for the development of increasingly more realistic models and more powerful multi-scale simulation methods as they are being applied nowadays in the materials science community. We address in particular the development and application of a broad range of experimental and computational methods to describe the electronic, optical, and magnetic properties of matter, including nonadiabatic ultrafast phenomena and non-equilibrium dynamical processes at different size and time scales. This section plays a pivotal role in the whole structure of the *Handbook of Materials Modeling* by providing the fundamentals of the methodologies used in the two volumes. In particular, it has close ties with this volume's sections on atomistic simulations, materials informatics, molecular dynamics-thermodynamics, and magnetism-spintronics. Furthermore, it offers complementary information to many of the sections in the second volume of the Handbook that focus on applications to demonstrate and expand the capabilities of current models and simulation methods, in particular the one dealing with photovoltaics from first-principles.

At the simplest level of treating the many-electron problem, the Hartree-Fock theory (HF) is obtained by considering the wave function to be a single Slater determinant. In this way the N-body problem is reduced to N one-body

problems with a self-consistent requirement due to the dependence of the HF effective potential on the wave functions. By the variational theorem, the HF total energy is a variational upper bound of the ground-state energy for a particular symmetry. The HF eigenvalues may be used as estimates of the true excitation energies. The HF theory is far from accurate because the wave function of a system in general cannot be written as a single determinant for the ground state and Koopman's theorem is in general a poor approximation. On the one hand, a rigorous method is to employ the many-particle Green's function approach in which the poles of the one-particle and two-particle Green's function give the quasiparticle energies and optical transition energies, respectively. On the other hand, within the density-functional theory (DFT), the ground-state energy of an interacting system of electrons in an external potential can be written as a functional of the ground-state electronic density. When comparing to conventional quantum chemistry methods, this approach is particularly appealing since it does not rely on the complete knowledge of the N-electron wave function but only on the electronic density. However, although the theory is exact, the energy functional contains an unknown quantity called the exchange-correlation energy, $E_{xc}[n]$, that has to be approximated in practical implementations. For excited-state properties, the DFT has been extended to a time-dependent formalism (TDDFT) and applied with success to the calculations of optical properties and excited-state dynamics of finite and extended systems. However, the commonly used adiabatic functionals in DFT and TDDFT encounter problems when studying correlated materials and spectroscopic properties of extended systems and excitation energies in molecules. Of course, when going to more complex systems, the key to describe and model most physical effects in chemistry and materials is not achieving "chemical accuracy" but being able to theoretically describe and capture the relevant physical-chemical processes.

When describing light-matter interactions, those theoretical tools have a certain range of validity as they often treat electromagnetic radiation and matter on a different level of approximation. Photons are usually only treated as an external perturbation that probes matter without any further influence and vice versa. The standard theoretical modeling can thus be insufficient when photon and matter degrees-of-freedom become equally important (as in the case of polariton condensates). Indeed, these matter-only approaches can be viewed as approximations to a density functional or Green's function formulation of nonrelativistic quantum electrodynamics (QED). In my group we have recently introduced a novel quantum-electrodynamical density-functional approach (QEDFT) to describe such complex dynamics of interacting electrons, photons, and phonon systems all on the same theoretical footing (Ruggenthaler et al. 2018; Flick et al. 2015, 2017). QEDFT is a novel framework to deal with electron-photon interactions from first-principles (opening up the fields of cavity QED-chemistry and QED materials). Hybridizing light strongly with the electronic structure of the system, novel effects appear providing a promising route for new design of material, for example, Floquet engineering (quantum topological matter), valley Hall effect in 2D materials, chiral plasmonics, or phonon-driven spin-magneto valleytronics. Moreover, hidden

aspects of photon-matter interaction can be revealed with new tailored spectroscopic tools (Ruggenthaler et al. 2018; Flick et al. 2015, 2017).

## 2 Contributed Chapters to "Electronic Structure of Materials by Ab Initio Methods"

The carefully chosen contributions presented here handle fundamental aspects of the different frameworks to deal with static and dynamical properties of many-body interesting systems (from molecules to solids). We touch up on basic ideas coming from traditional quantum chemistry methods, condensed matter approaches based on different flavors of many-body perturbation theory as well as static and time-dependent density (and density matrix) functional approaches. Thus, the articles collected in this section cover a wide and impressive range of topics in the fields of chemistry, physics, and material science.

In the next 12 chapters (and in some of the 4 plenary presentations), we provide an in-depth discussion of the basic concepts of density-functional theory and the development of exchange and correlation functionals (▶ Chap. 11, "Recent Developments in Density Functional Approximations," L. Li and K. Burke), which are able to cope with strong correlations, charge transfer phenomena (▶ Chap. 12, "Charge Transfer in Molecular Materials," T. Zhu, T. Van Voorhis, P. de Silva) and dispersion (van der Waals) (▶ Chap. 13, "Van der Waals Interactions in Material Modelling," J. Hermann and A. Tkatchenko) forces within a DFT formalism and beyond, both in the static as well as in time-dependent domains (touching on the efficient ab initio modeling of pump probe spectroscopies (▶ Chap. 14, "Pump-Probe Photoelectron Spectra," U. De Giovannini)). After having presented the approach to solve the many-body problem from a density-functional perspective, we change gear and address the recent exciting developments in correlated many-body techniques based on Green's functions methods, including new developments on the GW+BSE framework for fermions (▶ Chap. 15, "Modeling Excited States of Confined Systems," L. Hung and S. Öğüt) and coupled fermion-boson systems (▶ Chaps. 16, "Many-Body Calculations of Plasmon and Phonon Satellites in Angle-Resolved Photoelectron Spectra Using the Cumulant Expansion Approach," F. Caruso, C. Verdi, F. Giustino, and ▶ 17, "Non-equilibrium Green's Functions for Coupled Fermion-Boson Systems," D. Karlsson and R. van Leeuwen), non-equilibrium dynamical mean field approaches (▶ Chap. 18, "Non-equilibrium Dynamical Mean-Field Theory," M. Eckstein), quantum (▶ Chap. 19, "Correlations and Effective Interactions from First Principles Using Quantum Monte Carlo," L. Wagner), and diagrammatic Monte Carlo approaches (▶ Chap. 20, "Diagrammatic Monte Carlo and GW Approximation for Jellium and Hydrogen Chain," K. Van Houcke, IS Tupitsyn and NV. Prokof'ev) as well as coupled cluster and quantum chemistry schemes (▶ Chap. 21, "Coupled Cluster and Quantum Chemistry Schemes for Solids," A. Grüneis). This section ends with a contribution on how those different electronic structure methods could be combined with optimal control theory in order

to design protocols that allow one to achieve specific predefined targets (▶ Chap. 22, "Optimal Control Theory for Electronic Structure Methods," A. Castro), i.e., bring a system into a specified electronic state, maximize a higher order response function, improve the yield of a chemical reaction, etc. Those methods are just several selected from a lot of examples, given to illustrate the current status in ab initio calculations. Yet work needs to be done toward the inclusion of higher order vertex effects into the many-body description of electron-electron and electron-phonon interactions in order to have a more general ab initio theory of the spectra of weak and strongly correlated electronic systems.

## 3      Final Concluding Remarks

In conclusion, this section is a journey through some of the recent advances in solving the many-body electronic problem in complex materials, which have been enabled by major theoretical and methodological developments. Some of the methods and ideas discussed are very recent and will clearly evolve in the next years. Also, I want to mention here three other schemes that have emerged as powerful tools to deal with correlated many-body systems that have not been addressed by any of the chapters in this section, just to name a few: density-matrix renormalization group (Schollwöck 2011) and embedding methods (Wouters et al. 2016), stochastic method for DFT (Cytter et al. 2018) and many-body schemes (Neuhauser et al. 2014), auxiliary field quantum Monte Carlo (Motta and Zhang 2017), (the recent advances on this topic are reviewed in ▶ Chap. 7, "Ab Initio Electronic Structure Calculations by Auxiliary-Field Quantum Monte Carlo" by S. Zhang in this volume) machine learning for many-body problems (Carleo and Troyer 2017), quantum electrodynamical density-functional theory (Ruggenthaler et al. 2018; Flick et al. 2015). This is just an indication of the dynamism and wealth of ideas and methodologies being explored in our community to enable to tackle the many-body problem in realistic interacting quantum materials. This is clearly an exciting field of research.

Especially in this field, we have been witness of a real knowledge transfer from academia to society in the last years, with the rise and establishment of many start-up companies based on such accumulated knowledge (we are clearly entering the age of big data analytics in many fields arounds us, and computational materials science has joined the wave since the start). In this context, the whole Handbook and in particular this fundamental section on electronic structure will play an important role in further enabling and strengthening such synergy between fundamental research and society.

I don't want to take more space and much rather would like to leave you enjoying the reading of the next articles in which many experts have laid out the fundamentals of the electronic structure methods being in use now, which will shape the field of chemistry and condensed matter physics in the next years. I hope you will find those contributions interesting and relevant for your scientific work or scientific curiosity.

# References

Carleo G, Troyer M (2017) Solving the quantum many-body problem with artificial neural networks. Science 355:602

Cytter Y, Rabani E, Neuhauser D, Baer R (2018) Stochastic density functional theory at finite temperatures. Phys Rev B 97:115207

Flick J, Ruggenthaler M, Appel H, Rubio A (2015) Kohn-Sham approach to quantum electro-dynamical density functional theory: exact time-dependent effective potentials in real space. PNAS 112(15):285

Flick J, Ruggenthaler M, Appel H, Rubio A (2017) Atoms and molecules in cavities: from weak to strong coupling in QED chemistry. PNAS 114:3026

Motta M, Zhang S (2017) Computation of ground-state properties in molecular systems: back-propagation with auxiliary-field quantum Monte Carlo. J Chem Theory Comput 13:5367c

Neuhauser D, Gao Y, Arntsen C, Karshenas C, Rabani E, Baer R (2014) Breaking the theoretical scaling limit for predicting quasiparticle energies: the stochastic GW approach. Phys Rev Lett 113:076402

Ruggenthaler M, Tancogne-Dejean N, Flick J, Appel H, Rubio A (2018) From a quantum-electrodynamical light-matter description to novel spectroscopies. Nat Chem Rev 2, 0118

Schollwöck U (2011) The density-matrix renormalization group in the age of matrix product states. Ann Phys 326:96

Wouters S, Jiménez-Hoyos CA, Sun Q, Kin-Lic Chan G (2016) A practical guide to density matrix embedding theory in quantum chemistry. J Chem Theory Comput 12:2706

# Recent Developments in Density Functional Approximations

# 11

Li Li and Kieron Burke

## Contents

**Abstract**

We survey some of the standard approximations used in density functional calculations, most of which are at least 20 years old, and some new approaches that have been developed since.

L. Li

Department of Physics and Astronomy, University of California Irvine, Irvine, CA, USA
e-mail: li.li@uci.edu

K. Burke (✉)

Departments of Physics and of Chemistry, University of California Irvine, Irvine, CA, USA
e-mail: kieron@uci.edu

Table of acronyms in alphabetical order. References are given where they first appear in main text

|         | General terms                                               |
|---------|-------------------------------------------------------------|
|         | General terms                                               |
| DFT     | Density functional theory                                   |
| G3      | A standard database of molecular energy differences         |
| GGA     | Generalized gradient approximation                          |
| GKS     | Generalized Kohn-Sham                                       |
| HF      | Hartree-Fock                                                 |
| KS      | Kohn-Sham                                                    |
| LC20    | A database of 20 lattice constants of solids                |
| meta-GGA| An extension of GGA that uses the KS kinetic energy          |
| RPA     | Random-phase approximation                                   |
| S22     | A database of weakly bonded molecules                        |
| TDDFT   | Time-dependent density functional theory                    |
| XC      | Exchange-correlation                                         |
|         | Approximate functionals                                     |
| B3LYP   | The most popular functional in chemistry                    |
| B88     | An exchange GGA that is the most popular in chemistry        |
| LDA     | Local density approximation                                  |
| PBE     | A popular GGA in materials science                          |
| PW86    | A predecessor of PBE                                         |
| PW91    | The GGA that PBE is a simplification of                      |
| SCAN    | A recent meta-GGA                                            |

# 1    Introduction

Each year, at least 30,000 papers are published using density functional theory to perform electronic structure calculations (Pribram-Jones et al. 2015). Almost all solve the Kohn-Sham (KS) equations (Kohn and Sham 1965) self-consistently and use some approximation for the exchange-correlation (XC) energy as a functional of the (spin)-densities, $E_{XC}[n]$. This chapter surveys some of the more popular approximations. Some background is covered in Burke (2012), which is more focused on chemistry.

## 1.1    Commonly Used Approximations

The original approximation was suggested by Kohn and Sham themselves (Kohn and Sham 1965), namely, the local density approximation (LDA), in which the XC energy density at each point in the system is replaced by that of a uniform electron gas with the density at that point. The exchange contribution was first written by Bloch (1929), with correlation now well-known from quantum Monte

Carlo simulations of the uniform gas (Ceperley and Alder 1980), parametrized in simple formulas (Vosko et al. 1980; Perdew and Wang 1992). LDA is remarkably accurate for geometries but typically overbinds molecules by about 1 eV per bond, making it relatively useless for thermochemistry (Jones and Gunnarsson 1989).

The next step in complexity is the generalized gradient approximation (GGA), which creates an energy density using both the density and its gradient at each point (Burke et al. 1997). The basic concept and its first realization were given already in the pioneering work of Ma and Brueckner (1968). This was carefully refined in the work of Langreth, leading to the Langreth-Mehl functional (Langreth and Mehl 1981, 1983). Perdew pioneered the use of real-space cutoffs to create GGA's, leading to the PW86 functional (Perdew and Wang 1986). The highpoint of this detailed construction was the PW91 functional (Perdew 1991; Perdew et al. 1992; Burke et al. 1997), and in 1993, it was shown that they yield useful accuracy for binding energies, i.e., errors of about 6–10 kcal/mol (1 eV = 23 kcal/mol) (Gill et al. 1992). The most commonly used GGA in materials today is a simplification of the PW91 form called PBE (Perdew et al. 1996a), while in chemistry it is BLYP, with B88 exchange (Becke 1988) and Lee-Yang-Parr correlation (Lee et al. 1988). Both these are trained on noble gas atomic energies, yielding more accurate energies for those atoms (Elliott and Burke 2009; Burke et al. 2016).

The last standard step is to create a hybrid of GGA with the exact exchange energy from a Hartree-Fock calculation, by replacing a fraction, $a$, of the GGA exchange with the Hartree-Fock (HF) exchange, as first suggested by Becke (1993). This fraction is 20% in the famous B3LYP functional, (which stands for Becke, 3-parameter, Lee-Yang-Parr), the most commonly used approximation in chemistry today (Becke 1993; Lee et al. 1988; Vosko et al. 1980; Stephens et al. 1994). Its analog is PBE0 (Perdew et al. 1996b; Ernzerhof and Scuseria 1999) which uses 25% mixing.

At least 80% of all DFT calculations currently being performed use one of the approximations mentioned above. One can think of more sophisticated approximations using more ingredients of the density, as in Jacob's ladder (Perdew and Schmidt 2001). An important principle of progress in making density functional approximations is to ensure that (almost) everywhere, each level of approximation performs at least, as well as the previous level, and also improves some features, making the previous level obsolete (except to save computational time). One should avoid having different functionals for different purposes.

## 1.2   Beyond Ground-State DFT

What has been described above is generic ground-state DFT in the non-relativistic limit and for collinear magnetic fields. The single largest use of DFT beyond this domain is using time-dependent DFT (TDDFT) (Runge and Gross 1984; Maitra et al. 2004; Maitra 2016) in the linear response regime to extract electronic transition frequencies (Casida 1996; Petersilka et al. 1996). As much as 10% of all DFT publications include TDDFT estimates of excitations. Such calculations

almost all use the adiabatic approximation for the XC kernel and so are uniquely determined by the choice of ground-state approximation. Most such calculations are for molecules (Burke et al. 2005), as a nonlocal contribution is needed to yield details of bulk semiconductor spectra correctly (such as exciton binding energies), and that contribution is missing from the commonly used functionals mentioned above (Onida et al. 2002; Martin et al. 2016).

Other extensions include magnetic DFT for non-collinear fields, relativistic DFT (Engel and Dreizler 2011) and even QED (Flick et al. 2015), DFT at finite temperatures using the Mermin theorem (Smith et al. 2018), coupling between nuclei and electrons (Requist and Gross 2016), and so on. Again, standard ground-state approximations are used unless a particular effect requires going beyond these, such as extracting double excitations (Maitra et al. 2004).

## 2  Recent Developments

This section is devoted to developments over the past 20 years.

### 2.1  RPA-Type Functionals

The famous adiabatic-connection fluctuation-dissipation formula of DFT yields $E_{XC}$ in terms of the a frequency and coupling constant and spatial integral over the density-density response function (Langreth and Perdew 1975; Harris and Jones 1974; Gunnarsson and Lundqvist 1976). This can be extracted directly from the KS response function, constructed from the occupied and unoccupied orbitals, and the XC kernel of TDDFT (Gross and Kohn 1985). Ignoring the kernel yields RPA, also known as direct random-phase approximation (RPA) (Furche 2001; Langreth and Perdew 1977, 1975; Chen et al. 2017) or TD-Hartree. This scheme thus yields a fifth-rung (in Jacob's ladder) approximation that can be costly to evaluate, but the relative burden is always decreasing (Furche 2008; Eshuis et al. 2010). Direct RPA overcorrelates systems, because it includes only "bubble" diagrams in the many-body expansion of the energy, and misses other contributions at higher-order that reduce correlation. It also has difficulties with self-interaction, because it yields finite correlation energies even for only one electron. These two effects yield inaccuracies in the dissociation energies of molecules. Recent progress has included various approximations to the XC kernel to yield improved energetics and computational cost savings. A very recent development, using the exact frequency-dependent exchange kernel, and a clever and physically motivated resummation of higher-orders, appears to overcome stability problems and even allows the binding curve of $N_2$ to be accurately calculated (Erhard et al. 2016; Burke 2016) (which is difficult even in coupled-cluster theory, due to the multireference nature of the wave function at large separations).

## 2.2   Weak Interactions

Tremendous progress has been made in the last two decades for including weak, van der Waals (vdW) interactions into DFT calculations. The standard functionals model covalent, ionic, and metallic bonds reasonably well but fail for longer, weaker bonds. Because of their semilocal nature, they cannot yield weak-binding that falls off as $R^6$, where $R$ is the separation between two atoms or molecules. Thus corrections must be added to the standard functionals to capture these effects. There are three main schemes for doing this.

In the domain of explicit nonlocal density functionals, there is the sequence of approximations originally developed by Langreth and Lundqvist and co-workers (Dion et al. 2004). These approximations are derived nonempirically, remarkably starting from contributions to the uniform gas correlation energy. The original was from 2004, and an improved parameterization was given in vdW2 (Klimeš et al. 2009). These explicit density functionals can be applied to all materials, from molecules to solids, but have been designed assuming the system has a gap (not a metal) (Berland et al. 2015).

At the extreme opposite end of the scale, there are the parametric schemes of Grimme and co-workers (Grimme 2006), commonly referred to as DFT-D. These are not explicit density functionals, but rather additional energies that include estimates of the $C_6$ (and higher-order) coefficients in the asymptotic expansion of the vdW energy between atoms, combined with a damping factor to keep the contributions finite as the separation reduces. Such schemes require empirical parameters for each atom for a given standard XC approximation but can yield highly accurate energy curves for small weakly bound molecules (Burns et al. 2017).

In between is the scheme developed by Tkatchenko and Scheffler, which requires only one parameter for a given XC functional and produces accurate add-on corrections to DFT energies (Tkatchenko and Scheffler 2009). This has been expanded to incorporate collective electrostatic effects so that metals and materials in many dimensions and on different length scales can be treated (Hermann et al. 2017). A final scheme is that begun by Becke and Johnson, which uses the dipole moment of the exchange hole to determine $C_6$ (and higher) coefficients (Johnson and Becke 2006).

Of course, more expensive treatments, such as RPA mentioned above, automatically include approximations to the vdW forces.

## 2.3   Meta-GGAs

The third rung of Jacob's ladder is the meta-GGA, which adds a new ingredient beyond that of the density and its gradient. This is most often chosen to be the (positive) kinetic energy density of the KS orbitals. The aim for a good meta-GGA is to aim for the accuracy of hybrids without the computational overhead of the exact exchange contribution. The cost of exact exchange is relatively manageable in molecular calculations with atom-centered basis functions but can often be

prohibitive when using periodic boundary conditions and plane-wave basis sets. Running a range-separated hybrid (see next section) can sometimes take 1000 times longer than a typical GGA in a materials calculation.

Perdew and co-workers and many others have spent several decades developing meta-GGAs, with many flawed attempts (Sun et al. 2015). But the most recent effort, called strongly constrained and appropriately normed semilocal density functional (SCAN), has passed many standard tests and appears very promising to join the pantheon of commonly used functionals. The G3 dataset (Curtiss et al. 2005) is a standard set of chemical bonds that LDA overbinds by about 3 eV (typically about 1 eV/bond). PBE reduces this to about 1 eV, while SCAN reduces this to about 1/4 eV. SCAN also has errors that are 2–3 times smaller than PBE on the S22 dataset (Jurecka et al. 2006) of weakly bonded systems. At the same time, SCAN reduces errors in lattice parameters on the LC20 data (Sun et al. 2011) set from about 0.05Å in PBE to about 0.01Å. SCAN also yields better water properties than PBE (Perdew et al. 2017). On the other hand, the underestimation of chemical barrier heights by PBE is only mildly improved, by about a factor of 30% (whereas hybrids are often 2–3 times better). Thus, for many properties, SCAN yields accuracies similar to hybrid functionals but at a fraction of the computational cost (for materials codes).

## 2.4    Range-Separated Hybrids

The theory behind range separation is an exact one, developed first by Andreas Savin (Toulouse et al. 2009; Savin 1996; Leininger et al. 1997). One simply writes the Coulomb repulsion as a sum of a short-ranged contribution (decaying more rapidly than the inverse of the separation) and a long-ranged contribution, which has no Coulomb singularity at zero separation. One can then include one contribution as an interaction in some generalization of the KS equations and have the redefined XC contribution accounting for the other. This is all formally exact, and exact XC functionals exist for such schemes (though they differ from their regular KS counterparts).

But a plethora of choices now await. The first is the length scale on which the range separation is performed, often denoted $1/\omega$. This is a continuous parameter, and since approximations will be made to the corresponding $E_{XC}$, the results are sensitive to it. Just like the fraction of exact exchange in global hybrids, there is always a temptation to adjust it.

Furthermore, the separation into long- and short-ranged contributions can occur for the exchange contribution (most common), the correlation, or both. Finally, one may wish to treat the short-ranged contribution with an approximate functional or, in other situations, the long-ranged contribution! For example, a vexing problem in TDDFT is to accurately calculate charge-transfer excitations of well-separated donor-acceptor complexes. Including long-ranged Fock exchange exactly works very well for this problem (Stein et al. 2009). On the other hand, the very successful HSE06 functional (Heyd et al. 2003) is a range-separated hybrid, in which the long-

ranged exchange is treated with an approximate functional, but the short-ranged exchange is treated exactly, to model screening in an extended insulator (Janesko et al. 2009). This particular hybrid typically yields accurate gaps for moderate-gap semiconductors and insulators (Perdew et al. 2017).

## 2.5   Gaps of Solids

A crucial failure of the standard approximations is their inability to predict gaps of semiconductors and insulators. The LDA underestimates the gap of bulk Si by a factor of 2 and makes Ge a metal, and GGAs fare little better. From very early on, a great strength of the GW method  (Aryasetiawan and Gunnarsson 1998) has been its ability to provide accurate and reliable gaps.

To understand this issue, it is important to first note that the KS gap of a periodic solid does *not* match the fundamental gap of the solid (Perdew et al. 1982; Perdew 1985). All indications suggest that in fact LDA and GGA yield reasonably accurate KS gaps (i.e., close to the KS gap found with the exact ground-state functional) (Grüning et al. 2006). But, unlike LDA and GGA, the exact functional allows access to the fundamental gap, which is just $I - A$, the difference between the ionization energy and electron affinity of a system. Consider a very large but finite cluster of material. One can then add and subtract an electron to find $I$ and $A$ and deduce the exact gap. In fact, modern methods exist for doing this in a periodic calculation (Stadele et al. 1999). But in LDA or GGA, the added electron or hole delocalizes over the entire system and, because of their lack of a derivative discontinuity, $I - A$ collapses to the HOMO-LUMO KS energy difference, i.e., the KS gap.

A great success of the past two decades has been the accurate calculation of moderate gaps using hybrid functionals such as HSE06 (Heyd et al. 2003). This is achieved by going to a generalized KS scheme (Seidl et al. 1996), in which the orbital-dependent part of the functional is treated as in HF theory, not pure KS theory (which would require treating it with optimized effective potential (OEP) methods (Kümmel and Kronik 2008)). By having an orbital-dependent potential, one can show that the generalized Kohn-Sham (GKS) gap of such a calculation does match $I - A$, and because of the orbital-dependence, an approximate derivative discontinuity is included. Thus the GKS gap in such a calculation is the approximate fundamental gap and is wider than the corresponding KS gap. This is how hybrid functionals and meta-GGAs yield wider and generally more accurate gaps than GGA's (Perdew et al. 2017).

## 3   Challenges and Hopes

Here we review some of the more depressing failures of our current approximations.

## 3.1    Strong Correlation

The failure of DFT with standard approximations for strongly correlated systems
has been known since before its invention, as HF has problems for $H_2$ when it is
stretched (Heitler and London 1927). The problem can be analyzed and related to
localization/delocalization errors of the standard approximations when integer (or
half-integer) numbers of electrons localize on different sites (Cohen et al. 2008).
This is often called static correlation in quantum chemistry and involves the KS gap
between two states becoming very small and the exact many-body wave function
becoming an (almost) equal mixture of two Slater determinants.

It is important to note that the difficulties lie only with the failure of approx-
imations under these circumstances, rather than the KS scheme itself. This can
be beautifully illustrated with the two-site Hubbard model, for which it is trivial
to construct the exact KS system, even when strongly correlated (Carrascal and
Ferrer 2012). Even in realistic cases (albeit in 1d), the KS equations for a strongly
correlated system always can be made to converge to the exact ground energy and
density if the exact XC functional is used (Wagner et al. 2013). But of course many
of the features of the KS system do not resemble those of the physical system under
such conditions (Carrascal et al. 2015). This point is often confused by practitioners
of many-body theory. The differences between KS response functions and the many-
body analogs for strongly correlated systems is not a signal that a density functional
approximation is failing to yield accurate energies for such systems.

But strong correlation in solids is even more difficult than static correlation
in molecules. To see this simply, consider chains of uniformly spaced H atoms.
As the spacing is increased, an electron localizes on each site. For $H_2$, the true
wave function combines two Slater determinants. But for $H_4$, there are four such
determinants, and a DFT calculation with, say LDA, will break symmetry into four
different solutions, one of which will have lowest energy (Wagner et al. 2014). As
the number of atoms in the chain grows, so does the number of nearly degenerate
solutions, separated by spin excitations of very small energy. In the thermodynamic
limit, these become infinite, and the usual quantum chemical starting point of a
single Slater determinant becomes hopeless (Qiu et al. 2017). Because this is such
an important problem, vast amounts of research have been performed studying this
limit, especially by the group of Weitao Yang (Zheng et al. 2011), but also by
Scuseria (Motta et al. 2017) and Becke (Johnson and Becke 2017).

## 3.2    The Role of Empiricism

The most practical systematic approach to the construction of density functionals
has that been championed by Perdew: combine exact conditions that are relevant
to a given level of approximation with appropriate norms, such as the uniform gas

or hydrogen atom, to create approximations of tremendous generality (Medvedev et al. 2017). A key aspect of this approach is that, as one rises on Jacob's ladder, each successive approximation works better than the previous one (or at least is no worse) under almost all circumstances (and the cases where it does not are usually very informative). Thus, for a given computational cost, there is a single (or at least, very similar) obvious choice that is rarely worse than using a lower rung. Use of appropriate norms may appear empirical, but it can be understood as choosing parameters in the approximations to capture limits that have not been fully derived as yet but in principle could be. The extremely successful B88 exchange functional (Becke 1988; Elliott and Burke 2009) can be viewed as incorporating appropriate norms (exchange energy of atoms).

On the other hand, the profusion of inexpensive computing resources has led to many databases with either experimental results or those of high-level computational chemistry against which new approximations can be tested. It has also led to empirical fitting of density functional approximations with many parameters, as championed by Truhlar and co-workers (Zhao and Truhlar 2008, 2006; Zhao et al. 2006). This approach typically produces more accurate approximations than those of Perdew et al. for the systems and properties fitted and for related systems and properties. But it does not yield single universal approximations that generically improve over previous steps on the ladder. Such approximations can fail badly when applied beyond their range of applicability.

An entirely new approach to functional approximation is to use machine-learning to learn from accurate data (Snyder et al. 2012; Rupp et al. 2012; Bartók et al. 2010). This differs from the earlier approaches, because it automatically includes highly nonlocal contributions, as captured for example, in the kernel which measures density separations by integrals over the entire system (Snyder et al. 2012; Li et al. 2016b; Vu et al. 2015; Snyder et al. 2013). Recent advances include the first KS-MD simulation with an ML-DFT approximation to bypass solving the KS equations (Brockherde et al. 2017) and accurate approximation of the full interacting functional (including XC) (Li et al. 2016a) even for strong correlation and even for extended systems (but so far, only in 1d, because of the cost of generating accurate data).

## 3.3 New Horizons

An entirely new arena for DFT which has grown immensely in the last decade or so is applications to warm dense matter (Smith et al. 2018), with temperatures significant on the electronic scale (about 100,000 K), but not so high that Thomas-Fermi theory (or classical behavior) dominates. Applications range from modeling planetary interiors to inertial confinement fusion. This field is so "hot" that even the input to thermal LDA, the XC energy of a uniform gas as a function of temperature, is only now being calculated at high accuracy (Groth et al. 2017).

# References

Aryasetiawan F, Gunnarsson O (1998) The *GW* method. Rep Prog Phys 61(3):237. http://stacks.iop.org/0034-4885/61/i=3/a=002

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403. http://link.aps.org/doi/10.1103/PhysRevLett.104.136403

Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38(6):3098–3100. https://doi.org/10.1103/PhysRevA.38.3098

Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. J Chem Phys 98(7):5648–5652. https://doi.org/10.1063/1.464913

Berland K, Cooper VR, Lee K, Schröder E, Thonhauser T, Hyldgaard P, Lundqvist BI (2015) Van der Waals forces in density functional theory: a review of the vdW-DF method. Rep Prog Phys 78(6):066501

Bloch F (1929) Über die quantenmechanik der elektronen in kristallgittern. Zeitschrift für Physik A Hadrons and Nuclei 52(7):555–600

Brockherde F, Vogt L, Li L, Tuckerman ME, Burke K, Mller KR (2017) Bypassing the Kohn-Sham equations with machine learning. Nature Commun. https://doi.org/10.1038/s41467-017-00839-3

Burke K (2012) Perspective on density functional theory. J Chem Phys 136. http://link.aip.org/link/doi/10.1063/1.4704546

Burke K (2016) Viewpoint: improving electronic structure calculations. Physics 9(108). https://physics.aps.org/articles/v9/108

Burke K, Perdew JP, Wang Y (1997) Derivation of a generalized gradient approximation: the PW91 density functional, Plenum, p 81. http://link.springer.com/chapter/10.1007%2F978-1-4899-0316-7_7

Burke K, Werschnik J, Gross EKU (2005) Time-dependent density functional theory: past, present, and future. J Chem Phys 123(6):062206. https://doi.org/10.1063/1.1904586

Burke K, Cancio A, Gould T, Pittalis S (2016) Locality of correlation in density functional theory. J Chem Phys 145(5):054112. https://doi.org/10.1063/1.4959126, http://scitation.aip.org/content/aip/journal/jcp/145/5/10.1063/1.4959126

Burns LA, Faver JC, Zheng Z, Marshall MS, Smith DG, Vanommeslaeghe K, MacKerell AD Jr, Merz KM Jr, Sherrill CD (2017) The biofragment database (BFDb): an open-data platform for computational chemistry analysis of noncovalent interactions. J Chem Phys 147(16):161727

Carrascal DJ, Ferrer J (2012) Exact Kohn-Sham eigenstates versus quasiparticles in simple models of strongly correlated electrons. Phys Rev B 85:045110. http://link.aps.org/doi/10.1103/PhysRevB.85.045110

Carrascal DJ, Ferrer J, Smith JC, Burke K (2015) The Hubbard dimer: a density functional case study of a many-body problem. J Phys Condens Matter 27(39):393001. http://stacks.iop.org/0953-8984/27/i=39/a=393001

Casida ME (1996) Time-dependent density functional response theory of molecular systems: theory, computational methods, and functionals. In: Seminario JM (ed) Recent developments and applications in density functional theory. Elsevier, Amsterdam

Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45:566

Chen GP, Voora VK, Agee MM, Balasubramani SG, Furche F (2017) Random-phase approximation methods. Ann Rev Phys Chem 68(1):421–445. https://doi.org/10.1146/annurev-physchem-040215-112308, pMID: 28301757

Cohen AJ, Mori-Sánchez P, Yang W (2008) Insights into current limitations of density functional theory. Science 321(5890):792–794

Curtiss LA, Redfern PC, Raghavachari K (2005) Assessment of gaussian-3 and density-functional theories on the g3/05 test set of experimental energies. J Chem Phys 123:124107

Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density functional for general geometries. Phys Rev Lett 92(24):246401. http://link.aps.org/doi/10.1103/PhysRevLett.92.246401

Elliott P, Burke K (2009) Non-empirical derivation of the parameter in the B88 exchange functional. Can J Chem Ecol 87(10):1485–1491. https://doi.org/10.1139/V09-095

Engel E, Dreizler RM (2011) Density functional theory: an advanced course. Springer, Berlin

Erhard J, Bleiziffer P, Görling A (2016) Power series approximation for the correlation kernel leading to Kohn-Sham methods combining accuracy, computational efficiency, and general applicability. Phys Rev Lett 117:143002. https://link.aps.org/doi/10.1103/PhysRevLett.117.143002

Ernzerhof M, Scuseria GE (1999) Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. J Chem Phys 110:5029

Eshuis H, Yarkony J, Furche F (2010) Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. The Journal of Chemical Physics 132(23):234114

Flick J, Ruggenthaler M, Appel H, Rubio A (2015) Kohn-Sham approach to quantum electrodynamical density-functional theory: exact time-dependent effective potentials in real space. Proc Nat Acad Sci 112(50):15285–15290

Furche F (2001) Molecular tests of the random phase approximation to the exchange-correlation energy functional. Phys Rev B 64:195120

Furche F (2008) Developing the random phase approximation into a practical post-Kohn–Sham correlation model. J Chem Phys 129(11):114105

Gill PMW, Johnson BG, Pople JA, Frisch MJ (1992) An investigation of the performance of a hybrid of hartree-fock and density functional theory. Int J Quantum Chem 44(S26):319–331. https://doi.org/10.1002/qua.560440828

Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem 27(15):1787–1799

Gross EKU, Kohn W (1985) Local density-functional theory of frequency-dependent linear response. Phys Rev Lett 55:2850

Groth S, Dornheim T, Sjostrom T, Malone FD, Foulkes WMC, Bonitz M (2017) Ab initio exchange-correlation free energy of the uniform electron gas at warm dense matter conditions. Phys Rev Lett 119:135001. https://link.aps.org/doi/10.1103/PhysRevLett.119.135001

Grüning M, Marini M, Rubio A (2006) Density functionals from many-body perturbation theory: the band gap for semiconductors and insulators. J Chem Phys 124:154108

Gunnarsson O, Lundqvist B (1976) Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. Phys Rev B 13:4274

Harris J, Jones R (1974) The surface energy of a bounded electron gas. J Phys F 4:1170

Heitler W, London F (1927) Interaction between neutral atoms and homopolar binding according to quantum mechanics. Z Physik 44:455

Hermann J, DiStasio RA, Tkatchenko A (2017) First-principles models for van der Waals interactions in molecules and materials: concepts, theory, and applications. Chem Rev 117(6):4714–4758. https://doi.org/10.1021/acs.chemrev.6b00446, pMID: 28272886

Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened coulomb potential. J Chem Phys 118(18):8207–8215. https://doi.org/10.1063/1.1564060

Janesko BG, Henderson TM, Scuseria GE (2009) Screened hybrid density functionals for solid-state chemistry and physics. Phys Chem Chem Phys 11(3):443–454

Johnson ER, Becke AD (2006) Van der waals interactions from the exchange hole dipole moment: application to bio-organic benchmark systems. Chem Phys Lett 432(4–6):600–603

Johnson ER, Becke AD (2017) Communication: DFT treatment of strong correlation in 3D transition-metal diatomics. J Chem Phys 146(21):211105

Jones R, Gunnarsson O (1989) The density functional formalism, its applications and prospects. Rev Mod Phys 61:689

Jurecka P, Sponer J, Cerny J, Hobza P (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. Phys Chem Chem Phys 8:1985–1993

Klimeš J, Bowler DR, Michaelides A (2009) Chemical accuracy for the Van der Waals density functional. J Phys Condens Matter 22(2):022201

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140(4A):A1133–A1138. http://link.aps.org/doi/10.1103/PhysRev.140.A1133

Kümmel S, Kronik L (2008) Orbital-dependent density functionals: theory and applications. Rev Mod Phys 80(1):3–60. https://doi.org/10.1103/RevModPhys.80.3

Langreth D, Mehl M (1981) Easily implementable nonlocal exchange-correlation enery functional. Phys Rev Lett 47:446. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.47.446

Langreth D, Mehl M (1983) Beyond the local-density approximation in calculations of ground-state electronic properties. Phys Rev B 28:1809. https://journals.aps.org/prb/abstract/10.1103/PhysRevB.28.1809

Langreth D, Perdew J (1975) The exchange-correlation energy of a metallic surface. Solid State Commun 17:1425

Langreth D, Perdew J (1977) Exchange-correlation energy of a metallic surface: wave-vector analysis. Phys Rev B 15:2884

Lee C, Yang W, Parr RG (1988) Development of the colle-salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37(2):785–789. http://link.aps.org/doi/10.1103/PhysRevB.37.785

Leininger T, Stoll H, Werner HJ, Savin A (1997) Combining long-range configuration interaction with short-range density functionals. Chem Phys Lett 275(3-4):151–160

Li L, Baker TE, White SR, Burke K (2016a) Pure density functional for strong correlation and the thermodynamic limit from machine learning. Phys Rev B 94:245129. http://link.aps.org/doi/10.1103/PhysRevB.94.245129

Li L, Snyder JC, Pelaschier IM, Huang J, Niranjan UN, Duncan P, Rupp M, Mller KR, Burke K (2016b) Understanding machine-learned density functionals. Int J Quantum Chem 116(11):819–833. https://doi.org/10.1002/qua.25040

Ma SK, Brueckner K (1968) Correlation energy of an electron gas with a slowly varying high density. Phys Rev 165:18

Maitra NT (2016) Perspective: Fundamental aspects of time-dependent density functional theory. J Chem Phys 144(22):220901. http://scitation.aip.org/content/aip/journal/jcp/144/22/10.1063/1.4953039

Maitra NT, Zhang F, Cave RJ, Burke K (2004) Double excitations within time-dependent density functional theory linear response. J Chem Phys 120(13):5932–5937. https://doi.org/10.1063/1.1651060

Martin RM, Reining L, Ceperley DM (2016) Interacting electrons. Cambridge University Press, Cambridge

Medvedev MG, Bushmarinov IS, Sun J, Perdew JP, Lyssenko KA (2017) Response to comment on density functional theory is straying from the path toward the exact functional. Science 356(6337):496–496

Motta M, Ceperley DM, Chan GKL, Gomez JA, Gull E, Guo S, Jiménez-Hoyos CA, Lan TN, Li J, Ma F, Millis AJ, Prokof'ev NV, Ray U, Scuseria GE, Sorella S, Stoudenmire EM, Sun Q, Tupitsyn IS, White SR, Zgid D, Zhang S (2017) Towards the solution of the many-electron problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods. Phys Rev X 7:031059. https://link.aps.org/doi/10.1103/PhysRevX.7.031059

Onida G, Reining L, Rubio A (2002) Electronic excitations: density-functional versus many-body green's-function approaches. Rev Mod Phys 74(2):601–659. https://doi.org/10.1103/RevModPhys.74.601

Perdew JP (1985) What do the Kohn-Sham orbitals mean? How do atoms dissociate? Plenum, Density functional methods in physics. Springer, Boston, MA pp 265–308. https://link.springer.com/chapter/10.1007/978-1-4757-0818-9_10

Perdew JP (1991) Electronic structure of solids '91, Ziesche P, Eschrig H (eds) (Berlin: Akademie-Verlag) p. 11 Perdew JP and Wang Y (1992) Phys Rev B 45(13):244

Perdew JP, Schmidt K (2001) Jacobs ladder of density functional approximations for the exchange-correlation energy. In: AIP conference proceedings, AIP, vol 577, pp 1–20

Perdew JP, Wang Y (1986) Accurate and simple density functional for the electronic exchange energy: generalized gradient approximation. Phys Rev B 33:8800

Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B 45(23):13244–13249. http://link.aps.org/doi/10.1103/PhysRevB.45.13244

Perdew JP, Parr RG, Levy M, Balduz JL (1982) Density-functional theory for fractional particle number: derivative discontinuities of the energy. Phys Rev Lett 49:1691–1694. http://link.aps.org/doi/10.1103/PhysRevLett.49.1691

Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. Phys Rev B 46:6671

Perdew JP, Burke K, Ernzerhof M (1996a) Generalized gradient approximation made simple. Phys Rev Lett 77(18):3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865, ibid 78:1396(E) (1997)

Perdew JP, Ernzerhof M, Burke K (1996b) Rationale for mixing exact exchange with density functional approximations. J Chem Phys 105(22):9982–9985. https://doi.org/10.1063/1.472933

Perdew JP, Yang W, Burke K, Yang Z, Gross EKU, Scheffler M, Scuseria GE, Henderson TM, Zhang IY, Ruzsinszky A, Peng H, Sun J (2017) Understanding band gaps of solids in generalized Kohn-Sham theory. Proc Nat Acad Sci. http://www.pnas.org/content/early/2017/02/28/1621352114.full

Petersilka M, Gossmann UJ, Gross EKU (1996) Excitation energies from time-dependent density-functional theory. Phys Rev Lett 76:1212

Pribram-Jones A, Gross DA, Burke K (2015) DFT: a theory full of holes? Ann Rev Phys Chem 66(1):283–304. http://www.annualreviews.org/doi/abs/10.1146/annurev-physchem-040214-121420

Qiu Y, Henderson TM, Zhao J, Scuseria GE (2017) Projected coupled cluster theory. J Chem Phys 147(6):064111

Requist R, Gross EKU (2016) Exact factorization-based density functional theory of electrons and nuclei. Phys Rev Lett 117(19):193001

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52(12):997. http://link.aps.org/doi/10.1103/PhysRevLett.52.997

Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA (2012) Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 108:058301. http://link.aps.org/doi/10.1103/PhysRevLett.108.058301

Savin A (1996) On degeneracy, near-degeneracy and density functional theory. Technical report, Louisiana State University, Baton Rouge

Seidl A, Görling A, Vogl P, Majewski JA, Levy M (1996) Generalized Kohn-Sham schemes and the band-gap problem. Phys Rev B 53:3764–3774. http://link.aps.org/doi/10.1103/PhysRevB.53.3764

Smith JC, Sagredo F, Burke K (2018) Warming up density functional theory. In: Frontiers of quantum chemistry. Wójcik MJ, Nakatsuji H, Kirtman B, Ozaki Y (eds) Springer Nature, Singapore, pp 249–271. https://doi.org/10.1007/978-981-10-5651-2_11

Snyder JC, Rupp M, Hansen K, Mueller KR, Burke K (2012) Finding density functionals with machine learning. Phys Rev Lett 108:253002

Snyder JC, Rupp M, Hansen K, Blooston L, Müller KR, Burke K (2013) Orbital-free bond breaking via machine learning. J Chem Phys 139(22):224104

Stadele M, Moukara M, Majewski JA, Vogl P, Görling A (1999) Exact exchange Kohn-Sham formalism applied to semiconductors. Phys Rev B 59:10031

Stein T, Kronik L, Baer R (2009) Reliable prediction of charge transfer excitations in molecular complexes using time-dependent density functional theory. J Amer Chem Soc 131(8): 2818–2820. http://pubs.acs.org/doi/abs/10.1021/ja8087482

Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. J Phys Chem 98:11623

Sun J, Marsman M, Csonka GI, Ruzsinszky A, Hao P, Kim YS, Kresse G, Perdew JP (2011) Self-consistent meta-generalized gradient approximation within the projector-augmented-wave method. Phys Rev B 84(3):035117

Sun J, Ruzsinszky A, Perdew JP (2015) Strongly constrained and appropriately normed semilocal density functional. Phys Rev Lett 115:036402. http://link.aps.org/doi/10.1103/PhysRevLett.115.036402

Tkatchenko A, Scheffler M (2009) Accurate molecular Van der Waals interactions from ground-state electron density and free-atom reference data. Phys Rev Lett 102:073005

Toulouse J, Gerber IC, Jansen G, Savin A, Ángyán JG (2009) Adiabatic-connection fluctuation-dissipation density-functional theory based on range separation. Phys Rev Lett 102:096404. https://link.aps.org/doi/10.1103/PhysRevLett.102.096404

Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can J Phys 58(8):1200–1211. http://www.nrcresearchpress.com/doi/abs/10.1139/p80-159

Vu K, Snyder JC, Li L, Rupp M, Chen BF, Khelif T, Mller KR, Burke K (2015) Understanding kernel ridge regression: common behaviors from simple functions to density functionals. Int J Quantum Chem 115(16):1115–1128. https://doi.org/10.1002/qua.24939

Wagner LO, Stoudenmire EM, Burke K, White SR (2013) Guaranteed convergence of the Kohn-Sham equations. Phys Rev Lett 111:093003. http://link.aps.org/doi/10.1103/PhysRevLett.111.093003

Wagner LO, Baker TE, Stoudenmire M E, Burke K, White SR (2014) Kohn-Sham calculations with the exact functional. Phys Rev B 90:045109. http://link.aps.org/doi/10.1103/PhysRevB.90.045109

Zhao Y, Truhlar DG (2006) A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. J Chem Phys 125(19):194101. https://doi.org/10.1063/1.2370993

Zhao Y, Truhlar D (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor Chem Accounts 120:215–241

Zhao Y, Schultz NE, Truhlar DG (2006) Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. J Chem Theory Comput 2(2):364–382

Zheng X, Cohen AJ, Mori-Sánchez P, Hu X, Yang W (2011) Improving band gap prediction in density functional theory from molecules to solids. Phys Rev Lett 107:026403. http://link.aps.org/doi/10.1103/PhysRevLett.107.026403

# Charge Transfer in Molecular Materials

**12**

Tianyu Zhu, Troy Van Voorhis, and Piotr de Silva

## Contents

### Abstract

Charge transport in materials has an impact on a wide range of devices – LEDs, PVs, batteries, fuel cells, circuits, and sensors all prominently exploit charge transfer characteristics of the underlying materials. As material design becomes more sophisticated, molecular components are playing a larger role in these

T. Zhu · T. Van Voorhis (✉)
Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: tyzhu@mit.edu; tvan@mit.edu

P. de Silva
Department of Energy Conversion and Storage, Technical University of Denmark, Kgs. Lyngby, Denmark
e-mail: pdes@dtu.dk

227

applications, so that an understanding of charge migration in molecular systems is increasingly relevant. Here, we present a concise review of the principles of charge transfer in molecular materials. After a brief summary of the key concepts of Marcus theory, we discuss the key molecular and material properties that influence charge transfer and how they can be accounted for. Using organic PV and LED materials as a case study, we illustrate how these concepts can be used to better understand the microscopic properties that underpin device function in real devices.

# 1      Introduction

Charge transfer in materials is a central step in a host of technologies. In photovoltaic cells, charge separation turns sunlight into electricity (Sariciftci et al. 1992). In batteries, long-lived charge transfer stores electrical energy (Goodenough and Park 2013). In fuel cells, electron transfer converts chemical energy into mechanical work (Wang et al. 2011). In LEDs, charge recombination turns electrical current into visible light (Nakamura et al. 1994). In $CO_2$ sequestration, charge transfer harnesses electrical energy to perform a chemical transformation (Appel et al. 2013).

In each of these technological settings, the molecular picture of charge transfer is increasingly important. In many cases, such as fuel cells, at least one of the active phases is molecular in nature so that the need for a molecular description is obvious. In other cases, the motivation for a molecular picture is driven by the increasing success of molecular semiconductors. Whether one considers perovskite photovoltaics (Kojima et al. 2009) or organic LEDs (Tang and VanSlyke 1987), material discovery is consistently demonstrating that molecular materials are able to outperform traditional network solids in a variety of technological settings.

In this review, we discuss charge transfer in these types of molecular materials from a quantum chemical perspective. Whereas charge transport in inorganic semiconductors is typically described in terms of band theory (Kittel 2005) and delocalized charge carriers, in molecular systems, one typically employs Marcus theory (Marcus 1956; Hush 1961) in which localized charge carriers hop from one location to another. We begin by summarizing the basic principles of this Marcus picture, emphasizing its physical underpinnings and limitations. Next, we outline how modern quantum chemical tools can quantitatively simulate charge transfer within the Marcus framework. Finally, we use some examples involving organic solar cells and organic LEDs to illustrate the utility of these ideas and suggest directions for future research efforts.

# 2      Marcus Picture of Charge Transfer

*The Marcus picture describes electron (or hole) transfer from one localized center (the donor) to another localized center (the acceptor)* (Marcus 1956). Hence, by its very nature, the Marcus picture assumes that electrons localize on individual

centers. The resulting hopping-like picture of charge transport is thus fundamentally irreconcilable with the band theory of transport common in solid-state physics. Broadly speaking, the localized picture is expected to be appropriate in more disordered materials, because disorder drives quantum-mechanical localization (Anderson 1958). Molecular systems are only held together by weak intermolecular forces, making them highly defective compared to inorganic network solids. The resulting rugged energy landscape tends to localize charges, making the Marcus picture the default approach in molecular materials (Coropceanu et al. 2007).

The two electronic states in which the electron is localized on one center or the other are called *diabatic states* (Mead and Truhlar 1982; Van Voorhis et al. 2010). Typically, one envisions localization on a single molecule, although submolecular localization (e.g., within a polymer) or super-molecular localization (e.g., to a single aggregate) is also possible. In any case, charge transfer between the centers is driven by fluctuations in the energies of the diabatic states. Again, molecular materials differ from their inorganic counterparts in this respect. Whereas inorganic network solids tend to be inflexible, organic materials are comparatively soft and deformable, leading to relatively large, dynamic fluctuations in the energies of the diabatic states. These fluctuations in turn drive efficient charge transfer. As illustrated in Fig. 1a, these fluctuations can arise both from intramolecular reorganization (e.g., changes in bond lengths or torsion angles) and intermolecular reorganization (e.g., molecular reorientation or polarization), but the result is the same in either case: dynamic changes in the donor and acceptor energies occasionally lead to configurations where the two states are degenerate and charge transfer can occur rapidly. The rate of charge transfer is thus partly determined by the probability of obtaining one of these rare degenerate configurations.

For any given system, identifying the geometric changes that govern these fluctuations (the so-called reaction coordinate) is a task unto itself (Marcus 1993). However, the structural changes that accompany electron transfer tend to be rather system-specific – the reaction coordinate for electrochemistry in solution (Blumberger and Sprik 2006), for example, might be very different than for hole transfer in a polymer (Brédas et al. 2004). This variability tends to complicate the discussion of electron transfer in different systems that are otherwise physically and mathematically analogous. As a result, it is common to instead use the energy gap, $q \equiv \Delta E = E_{Donor} - E_{Acceptor}$, as a universal reaction coordinate (Warshel 1982; Tachiya 1993). Clearly this gap will be negative for configurations where the electron would prefer to be on the donor (the "reactant") and positive for configurations that favor the acceptor (the "product") and so fulfills the key purpose of the reaction coordinate: it distinguishes the product from the reactant.

Now, consider the probability, $P(\Delta E)$, of observing a given energy gap in our system. Roughly speaking, the rate of charge transfer will be given by (Marcus 1964):

$$k_{CT} \propto \left| \left\langle \Psi_{Donor} \left| \widehat{H} \right| \Psi_{Acceptor} \right\rangle \right|^2 P(\Delta E = 0) \tag{1}$$
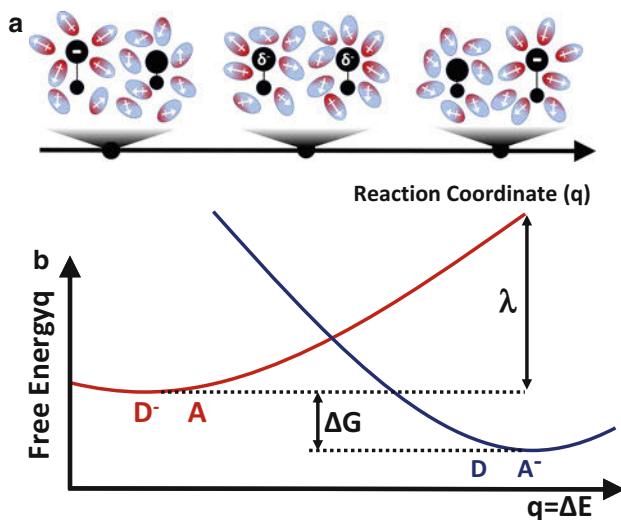
**Fig. 1** (**a**) The reaction coordinate in electron transfer involves molecular rearrangements and polarization that go from stabilizing charge on the "donor" (left) toward stabilizing charge on the acceptor (right). (**b**) The free energies of the system when the charge is on the donor (red) or acceptor (blue) resemble two parabolas that intersect at the transition state for charge transfer. Assuming the curves are exactly parabolic leads to the Marcus non-adiabatic rate expression

where the first term on the right-hand side is the electronic coupling between the donor and acceptor. One can derive Eq. 1 starting from Fermi's golden rule assuming nuclear motion is classical (Barbara et al. 1996), but the physical justification is clear. Charge transfer between two localized states is governed by the product of two factors: (1) the probability of attaining a configuration where donor and acceptor are degenerate ($P(\Delta E = 0)$) and (2) the probability of quantum-mechanical tunneling between donor and acceptor ($\propto \left| \langle \Psi_{\text{Donor}} \left| \widehat{H} \right| \Psi_{\text{Acceptor}} \rangle \right|^2$).

Visualization of the electronic coupling is challenging, but a qualitative picture of the probability is straightforward. If the electron is on the donor, we will tend to observe configurations in which $\Delta E$ will be negative. That is, $P_{donor}(\Delta E)$ will tend to be peaked at some negative value of $\Delta E$. Meanwhile, the opposite will be true of the acceptor: $P_{acceptor}(\Delta E)$ will tend to be peaked at some positive value of $\Delta E$. Thus, if we consider the free energy of each species (where $G_x(\Delta E) \equiv -kT \ln P_x(\Delta E)$), then we will obtain a picture similar to Fig. 1b: two free energies with minima at different points along the reaction coordinate that cross at the transition state ($\Delta E = 0$). Visually, these two curves resemble parabolas, and one typically makes the linear response approximation that these curves actually are parabolic, in which case the rate can be expressed in closed form:

$$k_{CT} \propto \left| \langle \Psi_{\text{Donor}} \left| \widehat{H} \right| \Psi_{\text{Acceptor}} \rangle \right|^2 e^{-\frac{(\Delta G + \lambda)^2}{4\lambda kT}} \qquad (2)$$

Equation 2 is the celebrated Marcus rate for non-adiabatic electron transfer. It expresses the rate in terms of three physical parameters: (1) the electronic coupling $V_{DA} \equiv \langle \Psi_{Donor} | \hat{H} | \Psi_{Acceptor} \rangle$, (2) the free energy difference ($\Delta G$) between donor and acceptor, and (3) the reorganization energy ($\lambda$) that measures the degree of structural relaxation that accompanies electron transfer.

As can be seen from the discussion above, there are a number of approximations built in to Eq. 2. There are therefore various extensions to the Marcus expression that can be considered:

- **Strong coupling**: By invoking Fermi's golden rule, one implicitly assumes that the electronic coupling, $V_{DA}$, is small. One can extend Eq. 2 to include higher-order terms in the coupling – effects like superexchange (Siddarth and Marcus 1993; Cheng et al. 2017) and even the adiabatic limit (Marcus and Sutin 1985; Jortner and Bixon 1988). In the context of materials, these higher-order terms eventually connect the Marcus picture to the traditional band picture of transport (Nan et al. 2009; Shuai et al. 2014).
- **Variable coupling**: In the above, we have assumed that the electronic coupling is a single number, whereas in reality, it will depend on molecular structure. Variation of the electronic coupling can influence the rate (Difley et al. 2010), and these variations can have a crucial impact in highly ordered molecular crystals (Troisi and Orlandi 2006a, b; Vehoff et al. 2010).
- **Beyond linear response**: The linear response approximation in Marcus theory can be justified heuristically using the central limit theorem, but for many systems, the free energy curves are markedly non-parabolic. In these cases, corrections for anharmonicity are available (Yeganeh and Ratner 2006), although less widely explored in molecular solids.

We will not dwell much on these extensions in what follows, in part because the extensions of Eq. 2 are typically system-specific. That is to say, whereas Eq. 2 applies universally to any system that satisfies a handful of physical approximations, the ways in which individual systems violate Eq. 2 are highly nonuniversal. Thus, in order to maintain a focused review, we will direct our attention toward computational realization of Eq. 2, leaving the discussion of extensions to other work.

## 3 Computational Realization

### 3.1 Calculations of Charge-Localized States and Electronic Couplings

Marcus theory provides a closed form expression for the charge transfer (CT) rate. Its practical use hinges on the possibility of calculating diabatic charge-localized states and electronic couplings. Standard electronic structure methods diagonalize some effective Hamiltonian; therefore, they generate adiabatic states

that are approximations to the true eigenstates of the Hamiltonian for the donor (A)-acceptor (B) system. In the diabatic basis, the Hamiltonian is not diagonal anymore, and the electronic coupling is simply its off-diagonal term:

$$V_{AB} = \left\langle \Psi_A \left| \widehat{H} \right| \Psi_B \right\rangle = \left\langle \Psi_{Donor} \left| \widehat{H} \right| \Psi_{Acceptor} \right\rangle \tag{3}$$

Since strict diabatization of the adiabatic basis is in general not possible (Mead and Truhlar 1982), a number of practical methods to generate charge-localized states have been developed. The practical concerns are computational feasibility and the cost of the diabatization procedure as well as accuracy of the resulting electronic couplings. The conceptually simplest method relies on modeling of the donor-acceptor complex as a two-level system (Newton 1991). The eigenvalues of the Hamiltonian are given by the expression:

$$E_{0,1} = \frac{1}{2} \left( E_A + E_B \pm \sqrt{(E_A - E_B)^2 + 4|V_{AB}|^2} \right) \tag{4}$$

In the case of a symmetric dimer ($E_A = E_B$), the expression simplifies, and the electronic coupling is just half of the energy gap between the adiabatic states:

$$|V_{AB}| = \frac{1}{2} (E_1 - E_0) \tag{5}$$

Therefore, for a symmetric donor-acceptor system, no explicit diabatization is needed, and the coupling can be computed just from the knowledge of the ground and first excited state energies. This can be simplified even further if the one-electron approximation is considered and adiabatic energies $E_0$, $E_1$ are replaced by orbital energies $\epsilon_{LUMO}$, $\epsilon_{LUMO+1}$ for electron and $\epsilon_{HOMO-1}$, $\epsilon_{HOMO}$ for hole transfer.

In a general case, one has to decide on the definition of a charge-localized state that will be further used in calculations of couplings and driving forces. Within the fragment orbital (FO) approach (Senthilkumar et al. 2003), diabatic states are defined by the orbitals of isolated molecules. In the simplest variant, the frozen orbitals of neutral molecules are considered, so only relevant frontier orbitals participate effectively in the electron transfer process, and the coupling element is

$$V_{AB}^{FO} = \left\langle \phi_A^{HOMO/LUMO} \left| \widehat{h} \right| \phi_B^{HOMO/LUMO} \right\rangle \tag{6}$$

where $\widehat{h}$ is the effective one-electron Hamiltonian constructed from the fragment orbitals. HOMOs or LUMOs of the A/B fragments are considered for hole and electron transfer, respectively. Such approach is computationally very efficient; however, it neglects the effect of relaxation of orbitals upon charging/discharging and mutual polarization of fragments. Several flavors of the FO method have been proposed to address these shortcomings (Schober et al. 2016), which still rely on separate calculations for fragments, but take account of the correct number of electrons in the construction of the reference states and the Hamiltonian.

More accurate calculations of diabatic states should fully take into account the mutual interactions of the donor and acceptor fragments in their proper oxidation states. Since minimization of the total energy leads to the adiabatic ground-state wave function, one needs to impose some sort of a constraint that enforces an excess electron or a hole to localize on the desired fragment. This is achieved in constrained density functional theory (CDFT) (Wu and Van Voorhis 2005, 2006; Kaduk et al. 2012) through minimization of the total energy functional under an additional constraint that the charge densities of fragments integrate to the predefined numbers of electrons. Since there is no unique way to partition the electron density into fragments, a weight function $w(r)$ is introduced to define a charge constraint on fragment A:

$$N_A = \int w_A\left(r\right)\rho\left(r\right)dr \tag{7}$$

While there is much freedom in the definition of the weight function, in the majority of implementations, it is based on Becke grid weights (Becke 1988) or Hirshfeld density partitioning (Hirshfeld 1977). Despite this ambiguity, the electronic couplings computed with CDFT states show only limited sensitivity to the choice of the weight function (Goldey et al. 2017).

In CDFT, the energy is minimized subject to the constraint by finding a stationary point of the following Lagrangian:

$$W\left[\rho, V_A\right] = E\left[\rho\right] + V_A\left(\int w_A\left(r\right)\rho\left(r\right)dr - N_A\right) \tag{8}$$

This leads to Kohn-Sham equations that need to be solved to obtain a KS charge-localized state:

$$\left(-\frac{1}{2}\nabla^2 + v_H\left(r\right) + v_{xc}\left(r\right) + V_A w_A\left(r\right)\right)\phi_i\left(r\right) = \epsilon_i\phi_i\left(r\right) \tag{9}$$

where $v_H$ is the Hartree potential, $v_{xc}$ is the exchange-correlation potential, and $V_A w_A$ is the constraining potential that enforces the proper charge on fragment A. Since the optimal Lagrange multiplier $V_A$ is not known a priori, it has to be optimized together with the electron density. In practice, this is done by adding an extra optimization loop to enforce Eq. 7 at each SCF cycle. The resulting diabatic state is a Kohn-Sham determinant, which formally is only an auxiliary mathematical object used to represent the electron density. This poses a challenge for the calculation of electronic couplings (Eq. 3), which require knowledge of the actual wave functions of the diabatic states. Since diabatic states are eigenstates of the Hamiltonian with the constraining potential, the coupling can be rewritten in the following way (Wu and Van Voorhis 2006):

$$V_{AB} = \langle \Psi_A \, | \widehat{H} | \, \Psi_B \rangle = \frac{F_A + F_B}{2} \langle \Psi_A \mid \Psi_B \rangle - \frac{1}{2} \langle \Psi_A \, | V_A w_A \, (\boldsymbol{r}) + V_B w_B \, (\boldsymbol{r}) | \, \Psi_B \rangle \tag{10}$$

where $F_{A/B} = E_{A/B} + V_{A/B} N_{A/B}$ are energies of the diabatic states interacting with the constraining potential. In this form, the dependence of the electronic coupling on the wave functions is reduced only to the overlap and one-body terms. At this point, it is assumed that the Kohn-Sham determinant is a reasonable approximation to the wave functions, which turns out to work well in practice.

Another DFT-based approach to compute charge-localized states and electronic couplings takes advantage of splitting of the donor-acceptor complex into two distinct but interacting Kohn-Sham subsystems. In subsystem DFT (sDFT) (Cortona 1991; Jacob and Neugebauer 2014), the starting point is the partitioning of the electron density into fragments:

$$\rho_{AB} \, (\boldsymbol{r}) = \rho_A \, (\boldsymbol{r}) + \rho_B \, (\boldsymbol{r}) \tag{11}$$

Each of the subsystem densities is a ground-state solution of Kohn-Sham equations with constrained electron density (KSCED) (Wesolowski and Weber 1996), where interactions with the other subsystem are described through a local embedding potential:

$$\left( -\frac{1}{2} \nabla^2 + v_{\mathrm{KS}} \left[ \rho^{A/B} \right] (\boldsymbol{r}) + v_{emb} \, [\rho_A, \rho_B] \, (\boldsymbol{r}) \right) \phi_i^{A/B} \, (\boldsymbol{r}) = \epsilon_i \phi_i^{A/B} \, (\boldsymbol{r}) \tag{12}$$

where $v_{\mathrm{KS}}[\rho^{A/B}](\boldsymbol{r})$ is a regular Kohn-Sham potential for the subsystem $A/B$ and $v_{emb}[\rho_A, \rho_B](\boldsymbol{r})$ is the embedding potential which is a nonsymmetric bifunctional that couples the equations for both densities. In practice, the equations are solved one at a time while keeping the density of the other subsystem frozen until both densities converge. Such implementation of sDFT is closely related to frozen density embedding (FDE) theory (Wesolowski and Warshel 1993), and both terms are often used interchangeably in the literature.

The key feature of sDFT that allows for generating diabatic states is the non-uniqueness of the density partitioning in Eq. 11. In particular, the number of electrons assigned to subsystems A and B is completely arbitrary, so conceptually it is straightforward to construct a state with a predefined charge. It is not immediately clear why these charge-constrained states remain localized when the total energy of the donor-acceptor system is minimized. Since KSCED equations yield the exact density of the total system, the number of electrons assigned to subsystems A and B should not matter at convergence as the initial fragment densities would polarize and delocalize in order to add up to the correct one. Nevertheless, sDFT proved to be an efficient computational tool to generate charge-localized states (Pavanello and Neugebauer 2011; Pavanello et al. 2013) provided that initial densities are chosen as isolated fragments with proper charges and each state is expanded in an atomic basis set that is centered only on the given fragment's nuclei. Such choice biases

the solution toward the targeted diabatic states, and restricted basis prevents the undesired delocalization of their densities while letting them mutually polarize each other.

Computing of electronic couplings is problematic due to the use of Kohn-Sham determinants instead of wave functions of interacting electrons. In the mean-field approximation, the couplings can be rewritten as a functional of transition density and wave function overlaps:

$$V_{AB} = E\left[\rho^{AB}\right] det\left(\mathbf{S}^{AB}\right) \tag{13}$$

Here $\mathbf{S}^{AB}$ is a transition orbital overlap matrix with elements:

$$\mathbf{S}_{kl}^{AB} = \left\langle \phi_k^A \middle| \phi_l^B \right\rangle \tag{14}$$

and $\rho^{AB}(\mathbf{r})$ is the scaled transition density:

$$\rho^{AB}(\mathbf{r}) = \sum_{kl} \phi_k^A(\mathbf{r}) \left(S^{AB}\right)_{kl}^{-1} \phi_l^B(\mathbf{r}) \tag{15}$$

By assuming that Kohn-Sham orbitals can be used to approximate the diabatic states and using an approximate functional to evaluate $E[\rho^{AB}]$, electronic couplings between sDFT diabats are readily available through Eq. 13.

## 3.2    Incorporating Environmental Effects

Charge transfer reactions usually occur in condensed phases, and the environment can have a significant effect on electronic structure of reactants and the resulting charge transfer parameters. In fact, the environment's degrees of freedom play a crucial role in the CT process; their fluctuations make two diabatic states transiently isoenergetic, at which point the electron hop happens. On average, the environment stabilizes localized charges through polarization of its electron density, alignment of multipole moments, and geometric deformation (polaron formation). For these reasons, accounting for the presence of the environment in computational studies of CT is of fundamental importance. The methods for including embedding effects can be divided into explicit and implicit models, depending whether surrounding molecules are present explicitly in the simulations or are replaced by a dielectric continuous medium.

The Marcus rate expression involves two thermodynamic properties associated with the CT reaction, the driving force $\Delta G$ and the reorganization energy $\lambda$. The driving force is the difference between the free energies of the initial and final diabatic states, so the role of the environment model is to describe the stabilization of the charges after the environment's degrees of freedom have been equilibrated.

Calculations of the reorganization energies are more complicated as it is a non-equilibrium property, which requires calculations of the energy of one diabatic state in the equilibrium structure of the other. It is particularly challenging for the outer sphere component of $\lambda$, which is associated with the rearrangement of surrounding molecules, and its treatment depends on the type of the model used.

In modern implicit environment models (Cramer and Truhlar 1999; Tomasi et al. 2005), the molecule or the reactive complex is placed inside a cavity resembling its molecular shape. The cavity is constructed either as a union of van der Waals spheres or is enclosed inside an isosurface of the molecule's electronic density. All the space outside of the cavity is filled with a continuous medium characterized by its dielectric constant $\epsilon$, which is usually assumed to be frequency independent. The presence of the molecular system inside the cavity induces charges in the dielectric, which generate an effective reaction field interacting with the embedded electron density. The free energy of interactions with the environment can be calculated by solving the Poisson equation for the apparent charge on the surface of the cavity. The surface charge generates a Coulomb potential:

$$\phi_\sigma \left( \boldsymbol{r} \right) = \int \frac{\sigma \left( \boldsymbol{r}_s \right)}{\left| \boldsymbol{r} - \boldsymbol{r}_s \right|} \, d\boldsymbol{r}_s \tag{16}$$

which is then added to the Hamiltonian of the embedded molecule. In the self-consistent reaction field (SCRF) family of approaches, the equations for the electronic structure of the embedded system and the surface charge on the cavity are solved iteratively until convergence. The converged electron density and electrostatic potential are used to compute the electrostatic contribution to the embedding free energy:

$$\Delta G_{es} = \frac{1}{2} \int \phi_{reac} \left( \boldsymbol{r} \right) \rho \left( \boldsymbol{r} \right) d\boldsymbol{r} \tag{17}$$

where the reaction field $\phi_{reac}(\boldsymbol{r})$ is the difference between electrostatic potentials of a molecule in the cavity and vacuum. Apart from the electrostatic component, the free energy of cavitation and dispersion/repulsion interactions can be accounted for in implicit models. These contributions are usually highly parametrized to fit to experimental or other theoretical data.

The implicit models used most frequently in computational studies of charge transfer reactions in condensed phases are the polarizable continuum model (PCM) (Miertuš et al. 1981) in its several variants (Mennucci 2012), conductor-like screening model (COSMO) (Klamt and Schüürmann 1993; Klamt 2011), and the SMx family of models (Kelly et al. 2005; Marenich et al. 2007, 2009), where the latter are based on solution of the generalized Born instead of the Poisson equation. The common advantage of these methods is that they are computationally very efficient and account for the majority of important effects resulting from interaction with the environment. However, they rely on many fitted parameters, so transferability and accuracy are often an issue. Also, the specific interactions

between the molecules, e.g., hydrogen bonding, are not accounted for in continuum models. Another challenge of modeling of CT reactions with implicit environment is that they assume thermodynamic equilibrium, and there is no direct way to freeze environment's geometric relaxation in calculations of reorganization energies. Extensions of implicit models to non-equilibrium structures assume that the bulk dielectric constant has a slow and fast ($\epsilon_\infty$) component (Aguilar 2001), where the latter is associated with electronic polarization only (high-frequency component). Then, $\epsilon_\infty$ can be used in calculations of reorganization energies, which is a qualitative yet efficient approach.

While implicit models are efficient, they necessarily ignore the fact that the environment is composed of individual molecules. Conversely, in explicit environment models, the molecular system of primary interest is surrounded by a number of discrete molecules that constitute its environment. Various models differ by how interactions between different parts of the total system are treated. The conceptually simplest approach is to do calculations on the whole supersystem without singling out its individual components. Since a very large number of environment molecules often number, this approach can become impractical due to steep computational scaling of conventional electronic structure methods. Additionally, it may be difficult to disentangle the properties of the embedded system and the environment if their states begin to mix. This can be further exacerbated by errors of the computational method, e.g., delocalization error of approximate density functionals. These problems can be alleviated by the use of subsystem and embedding methods that perform calculations on individual components, while including the interactions with the others.

One of the fragment-based approaches that has been successfully used in CT simulations is subsystem DFT, which has been already discussed in the context of generating diabatic states. It is straightforward to generalize the density partitioning (Eq. 11) to an arbitrary number of subsystems:

$$\rho\left(\boldsymbol{r}\right) = \sum_{i}^{N_{sub}} \rho_i\left(\boldsymbol{r}\right) \tag{18}$$

where each subsystem density $\rho_i(\boldsymbol{r})$ typically represents one molecule. For each molecule, KSCED equations (Eq. 12) are solved, where the embedding potential is a functional of its density and the density of all the remaining fragments. Since the orbitals of the supersystem are never calculated, the method scales linearly with the number of solvent molecules and relatively large clusters can be simulated. It is also straightforward to calculate diabatic states of CT complexes interacting with their environment by assigning the desired number of electrons to two of the subsystems representing donor and acceptor molecules.

Further reduction of the computational cost of explicit models can be achieved by employing hybrid quantum mechanics/molecular mechanics (QM/MM) methods (Warshel and Levitt 1976; Field et al. 1990; Lin and Truhlar 2007). The supersystem is partitioned into the embedded fragment, usually the solute molecule or the donor-

acceptor complex, and the environment. The active subsystem is treated at the quantum-mechanical level (e.g., DFT), while the environment is described by a classical force field which does not explicitly account for the presence of electrons. The Hamiltonian for the total system is therefore partitioned into the QM, MM, and interaction terms:

$$H_{total} = H_{QM} + H_{MM} + H_{QM/MM} \qquad (19)$$

where $H_{QM}$ is the Hamiltonian accounting for interactions of nuclei and electrons of the embedded fragment, $H_{MM}$ accounts for interaction of MM atoms, and $H_{QM/MM}$ couples these two subsystems. The wave function of the quantum subsystem is optimized by minimizing the energy of the total system; therefore the polarization of the embedded subsystem by the environment is taken into account through the electrostatic term in the interaction Hamiltonian. The final interaction energy is given by the expression that accounts for coulomb interaction between solute and MM atoms represented by point charges and the additional empirical term representing van der Waals interactions:

$$
\begin{aligned}
E_{QM/MM} = & -\int \sum_{i \in MM} \frac{q_i}{|\boldsymbol{R}_i - \boldsymbol{r}'|} \rho\left(\boldsymbol{r}'\right) d\boldsymbol{r}' \\
& + \sum_{i \in MM} \sum_{j \in QM} \frac{q_i Z_j}{|\boldsymbol{R}_i - \boldsymbol{R}'|} + \sum_{i \in MM} \sum_{j \in QM} V_{vdW}\left(\boldsymbol{R}_i, \boldsymbol{R}_j\right)
\end{aligned}
\qquad (20)
$$

In simulations of charge transfer parameters, it is important to also include the polarization of the environment due to the presence of the donor-acceptor complex. The relaxation of the environment's electron density is the major contribution to the overall outer sphere reorganization energy and is chiefly responsible for the stabilization of charged diabatic states. Including this effect in QM/MM calculations requires the use of a polarizable force field in the MM part. A simple yet effective way of including polarization of MM atoms is based on Drude oscillators (Lamoureux and Roux 2003), i.e., fictitious charged and massless particles that are harmonically bound to nuclei. Drude particles are free to move around the atom to which they are anchored in response to the electric field of other atoms and Drude particles as well as the electron density of the QM region. The energy associated with Drude particles is

$$E_{Drude} = \sum_i \frac{1}{2} k_i^D |\boldsymbol{d}_i - \boldsymbol{r}_i|^2 + \sum_{ij} \frac{q_i^D q_j}{|\boldsymbol{d}_i - \boldsymbol{r}_i|} + \sum_{ij} \frac{q_i^D q_j^D}{|\boldsymbol{d}_i - \boldsymbol{d}_i|} \qquad (21)$$

Every Drude particle is characterized by its charge $q_i^D$ and spring constant $k_i^D$, which determine the atomic polarizability $\alpha_i = \frac{(q_i^D)^2}{k_i^D}$. In practice, often a constant value is assumed for the spring constant and the charges of Drude particles are fitted to reproduce molecular polarizability or atomic polarizabilities obtained through the

distributed multipole analysis (Stone 1981). Since the position of Drude particles and the electron density of the QM region depend on each other, calculations cycle between the optimization of the wave function and the position of Drude charges until self-consistency is reached.

### 3.3    Effects of Configurational and Energetic Disorder

Charge transfer reactions between molecular systems often occur in disordered and fluctuating environments like liquids, polymers, or glasses. Therefore, when simulating CT processes, it is usually not sufficient to consider only one configuration of the molecules surrounding the reaction center. This is especially true for charge-transporting molecular materials, where macroscopic properties like conductivity inherently depend on the morphology of the material. When a charge hops between molecular sites, the energetic landscape it experiences changes with the position and time. Incorporating these effects in molecular simulations requires a significant amount of computational resources but is often indispensable to understand macroscopic charge transport.

The overall disorder at any given instant of time has two primary sources. The dynamic disorder results from coupling of the electronic structure to the vibrations. Since vibrational modes of distinct molecules are not necessarily in phase, the molecules in the ensemble have different geometries and experience different local environments. This leads to a heterogeneity in CT parameters and resulting instantaneous CT transfer rates, which are a time-dependent property. If only dynamic disorder is present in the system, the distributions of these parameters sampled over time are identical for all the sites in the ensemble. In contrast, if there is static disorder present in the system, some of the sites have different time distributions of the CT parameters. The difference between the dynamic and static disorder is not always unambiguous as it depends on the time scale over which the properties are sampled. For ergodic systems, every molecule will eventually visit the entire phase space, so all the disorder in the ensemble is dynamic. However, in the context of CT reactions, the distinction between the two types of disorder is based on the time scale set by the charge transfer rate – motion that is slower than charge transfer effectively appears static to the transferring electron. In practice, it means that much of the disorder in molecular materials is inherently static, which may result from factors like sites being kinetically trapped in different local minima, presence of impurities, defects, phase boundaries, etc. The disorder of CT parameters can be further divided into the energetic and positional disorder. The former, which is also called the diagonal disorder, is manifested by the distribution of energies of diabatic states (site energies). The positional (off-diagonal) disorder gives rise to the heterogeneity of electronic couplings between different diabats. Static disorder typically reduces charge mobility of the material as the polaron encounters higher-energy barriers and often needs to move via percolation pathways. Large static disorder may also lead to charge trapping and subsequent chemical degradation of the material. In contrast, dynamic disorder is

fundamental for a charge hop to occur as the vibrations make the diabatic states transiently isoenergetic. Also, electronic couplings are very sensitive to the nuclear configuration so dynamic positional disorder affects how effectively the charge can be transferred between sites.

Atomistic computational studies of CT processes in a disordered material require the knowledge of its morphology under experimental conditions. Realistic configurations can be obtained from classical molecular dynamics (MD) simulations carried out in NVT or NPT ensembles. After initial equilibration, the snapshots of the MD trajectory are sampled at regular intervals, and these instantaneous configurations are used in subsequent QM calculations of CT parameters. In the hopping regime, only two molecular sites at a time are involved in charge transfer, but the environment has an important effect on the calculated values. Since nuclear configurations of the environment are sampled in the MD trajectory, explicit environment models can be directly used for every snapshot. In particular, polarizable QM/MM methods are well suited for such calculations (Difley et al. 2010) as they balance computational efficiency and inclusion of important physical effects. Sampling the energy landscape of the relevant diabatic states allows to compute the distribution of energy gaps, which can then be used to calculate free energy differences through the thermodynamic perturbation or thermodynamic integration methods (Frenkel and Smit 2002). In the linear response approximation, the energy gaps are assumed to be normally distributed, and the driving force for CT is expressed as the average energy gap in the reactant and product ensembles:

$$\Delta G = \frac{\langle \Delta E \rangle_A + \langle \Delta E \rangle_B}{2} \qquad (22)$$

while the reorganization energy is half of the difference between mean energy gaps

$$\lambda = \frac{\langle \Delta E \rangle_A - \langle \Delta E \rangle_B}{2} \qquad (23)$$

Analytical rate expressions enable calculation of the CT reaction rate between specific donor and acceptor sites. However, in a disordered material, there is a distribution of charge hopping rates which vary for different pairs of molecules involved in the CT reaction. To simulate the evolution of charge carriers, the material can be mapped to a lattice of molecular sites with assigned hopping rates between them. The mapping can be done based on atomistic simulations, where every molecule in an MD snapshot is represented by a lattice site and Marcus rates are calculated for every pair of nearest neighbors. Such approach reflects the realistic morphology of a material but is computationally demanding due to the large number of CT parameters that need to be computed. A simpler method that enables to study charge transport through a disordered medium is to assume a distribution of CT parameters and draw from this distribution to assign hopping rates based on some rate expression. In the Gaussian disorder model (GDM) (Bässler 1993), site energies $\epsilon$ are assumed to be normally distributed:

$$\rho\left(\epsilon\right) = \left(\sqrt{2\pi\sigma^2}\right)^{-1} exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \tag{24}$$

and the hopping rate between neighboring sites $i$ and $j$ is of the Miller-Abrahams type (Miller and Abrahams 1960):

$$v_{ij} = v_0\, exp\left(-2\gamma r_{ij}\right) \begin{cases} exp\left(-\frac{\epsilon_j - \epsilon_i}{kT}\right), & \epsilon_j > \epsilon_i \\ 1, & \epsilon_j < \epsilon_i \end{cases} \tag{25}$$

where the parameters $v_0$ and $\gamma$ are the frequency factor and the inverse localization radius of the charge wavefunction. Extensions of GDM (Pasveer et al. 2005; Bouhassoune et al. 2009; Kordt and Andrienko 2016) have been developed to account for effects like correlations of site energies or finite charge-carrier density. Evolution of a charge carrier in terms of site occupations $\{p_i\}$ is given by the following master equation:

$$\frac{\partial p_i}{\partial t} = \sum_j p_j \omega_{ji} - \sum_j p_i \omega_{ij} \tag{26}$$

In a general case, Eq. 26 cannot be solved analytically. Instead, charge dynamics can be directly simulated with the kinetic Monte Carlo (KMC) method (Bortz et al. 1975), which propagates the carrier by constructing a Markov chain based on predefined probabilities of transitions between different states (site occupations). Averaging of independent KMC trajectories allows to calculate macroscopic properties like charge mobility, diffusivity, charge density, and current (Kwiatkowski et al. 2008).

## 4 Computational Results

To illustrate how quantum chemistry methods can help with understanding charge transfer processes, we focus on their applications to organic semiconductors (OSCs). OSC materials hold promise for photovoltaic devices due to their low manufacture cost and for light-emitting diode devices because of their great optical properties and flexibility. In order to improve the efficiency of organic photovoltaics (OPVs) and organic light-emitting diodes (OLEDs), it is crucial to understand processes such as charge separation at the interface, charge recombination, charge transfer, and exciton diffusion in these devices. Therefore, we review some of the recent work on modeling OSC devices, with an emphasis on our work using QM/MM simulation techniques.

## 4.1    Band Gaps

Band gaps are the most fundamental property for understanding charge transfer in OSC devices. Especially, the optimization of OPV materials relies on estimating their band gaps, for which experimental techniques are time-consuming and have limited accuracy (Hüfner 2003; Zahn et al. 2006). Estimating band gaps by running DFT calculations on single molecules is inaccurate because it ignores the disordered environment in OSC devices. As discussed in Sect. 3.2, to incorporate the disordered and polarizable effects of the environment, it is necessary to use implicit solvent model (e.g., PCM) or explicit solvent model (e.g., QM/MM).

Recently, Kohn et al. (2017) modeled the PPV-PCBM and P3HT-PCBM OPV devices and compared the accuracy of different methods on bulk OSC band gaps. The authors first performed an MD simulation using OPLS force field (Jorgensen et al. 1996) to sample the bulk structures of PPV, P3HT, and PCBM. DFT calculations were then employed to compute ionization potential (IP) and electron affinity (EA) of the chosen OSC molecule from MD snapshots. The environment was treated in three different ways: as absent (vacuum), as point charges (non-polarizable QM/MM), and as point charges with corresponding Drude particles (polarizable QM/MM). The authors showed that non-polarizable QM/MM improved the estimation of both IP and EA values of PCBM by 0.5 eV over the vacuum but did not correct IP/EA values of P3HT and PPV. Adding Drude particles in polarizable QM/MM significantly improved the band gaps of all three OSC materials: the EA values of PCBM, PPV, and P3HT were closer to experimental values by 0.8 eV, 0.4 eV, and 0.9 eV, respectively. It was thus suggested that inclusion of environmental polarizability is crucial for obtaining quantitative description of OSC band gaps. This result is consistent with many earlier studies on OPVs and OLEDs (Difley et al. 2010; Yost et al. 2011, 2014; Yost and Van Voorhis 2013).

A later study by de Silva and Van Voorhis (2018) applied the same methods to an OLED emission layer consisting of PIC-TRZ2 and mCP. The authors found it is important to include polarizable environment while computing the band gaps of PIC-TRZ2 and mCP and showed this could be achieved by the polarizable QM/MM or PCM methods. However, this study found that although PCM can obtain quantitative mean values of band gaps, it failed to produce correct distributions of energy levels due to the lack of explicit environment. Therefore, the polarizable QM/MM method is a more accurate method to predict both mean values and distributions of band gaps of OSC materials.

In OSC devices, most of the important charge transfer processes happen at organic/organic interfaces, where the polarizable environment changes compared to the bulk. The change of polarization can affect the band gaps significantly. Such behavior is referred to band bending at the organic/organic interface in OPVs and is crucial for understanding charge separation mechanisms (Ishii et al. 1999). Quantum chemical simulations on several OPV devices have been performed to illustrate how band gaps change at the interface (Verlaak et al. 2009; Linares et al. 2010; Yost et al. 2011; Idé et al. 2013; Yost and Van Voorhis 2013). Figure 2 shows a recent study using QM/MM simulations on PPV-PCBM and P3HT-P3BM interfaces (Kohn
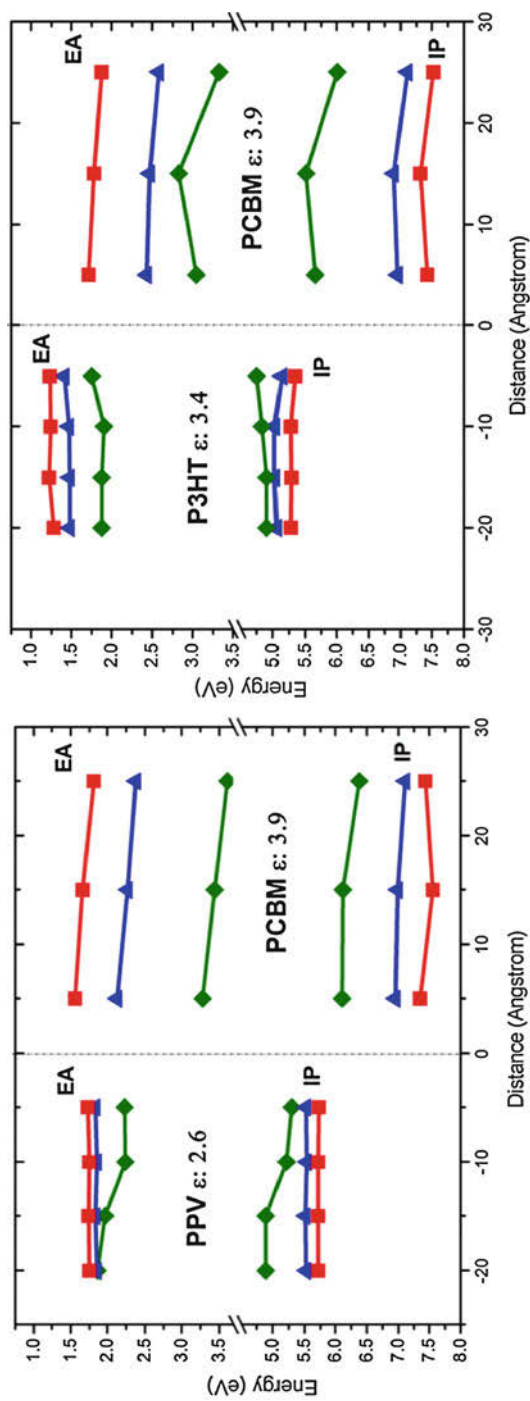
**Fig. 2** Band structures as a function of distance from the interface at the PPV-PCBM (left) and the P3HT-PCBM (right) interfaces with different treatment of the environment: as a vacuum (red), as static point charges (blue), or as static point charges with Drude particles (green) (Kohn et al. 2017)

et al. 2017). This study demonstrated polarization effects that result in notable band bending near the interface. For example, IP/EA values of PPV change by about 0.3 eV from the bulk to the interface suggested by the polarizable QM/MM method, which cannot be captured if the polarizable environment is ignored in simulations. Based on this result, the authors concluded that band bending at the PPV-PCBM interface promotes charge separation, while polarization effects inhibit charge separation in the P3HT-PCBM system.

## 4.2  Localization/Delocalization

In addition to band bending, another explanation for charge separation at the organic/organic interface in OPVs focuses on the role of delocalization of the charges in the charge-separated state (Nayak et al. 2013; Bässler and Köhler 2015). As donor materials in OPV devices are normally polymers, the hole state can delocalize easily over the donor, decreasing its Coulombic interaction with the electrons (McMahon et al. 2011; Castet et al. 2014). As a result, the hole-electron binding energy is reduced, and the charge separation at the interface becomes more efficient. Although the OPV acceptor materials are mostly fullerene-based, the delocalization of electrons over multiple fullerene units have also been reported (Jamieson et al. 2012; Savoie et al. 2014).

Kohn et al. (2017) have recently examined how the change of disorder and polarization at the interface affects delocalization in the hole electronic states in PPV-PCBM and P3HT-PCBM systems. Based on their QM/MM simulation results, they performed a normalized inverse participation ratio (NIPR) analysis, which counts the effective fraction of the polymer over which a hole state is spread. NIPR was computed as:

$$\text{NIPR} = \frac{1}{N} \frac{1}{\sum_{i \in \text{sites}} P_i^2} \tag{27}$$

where $P_i$ is the Mulliken population of the hole on each monomer $i$ and $N$ is the number of monomers in the polymer. The NIPR therefore indicates the extent of hole delocalization: for a polymer in which the hole is equally distributed over all $N$ monomers, NIPR $= 1$, and for a polymer in which the hole is only localized on one monomer, NIPR $= \frac{1}{N}$. Their NIPR results for hole states in PPV (14-mer) and P3HT (20-mer) are shown in Fig. 3, as a function of the distance from the interface. As can be seen, the NIPR for PPV does not change when the PPV chain is closer to the interface, suggesting the hole delocalization on PPV is insensitive to the PPV-PCBM interface. However, this is not the case for P3HT (which is more disordered than PPV): the hole distribution becomes obviously more localized when P3HT is closer to the P3HT-PCBM interface, indicated by the smaller NIPR values. The authors explained that this behavior arises from the interfacially induced disorder that disrupts the conjugation and causes a form of Anderson localization, which was also observed in other theoretical studies on P3HT (McMahon and Troisi
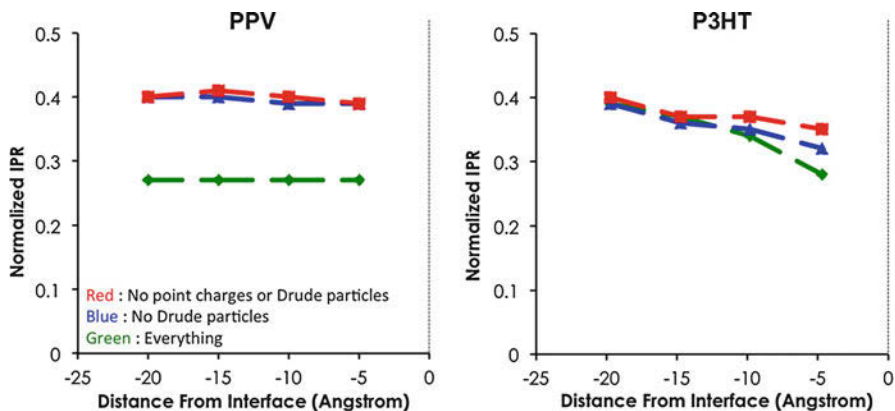
**Fig. 3** Normalized inverse participation ratios (NIPRs) for the hole states in the electron-donating polymers, PPV and P3HT, with different treatment of the environment: as a vacuum (red), as static point charges (blue), or as static point charges with Drude particles (green) (Kohn et al. 2017)

2010; Barford et al. 2010). The authors pointed out although the localization of hole states at the interface provides a driving force for charge separation, this effect is overpowered by the polarization of the electronic environment in the investigated systems. Therefore, they suggested considerations of electrostatic effects should take priority when designing new OPV devices.

Polarization effects from the environment show great importance in describing the hole delocalization. For PPV, the effect of static point charges added in the non-polarizable QM/MM is negligible because of PPV's less disordered structure. However, adding Drude particles in the polarizable QM/MM localizes the hole states in PPV by more than 30%. Such effects are smaller in P3HT: improved treatment of the polarizable environment increases the degree of localization but only reduces the hole size by about 15%. The authors thus concluded that the hole delocalization of disordered materials (such as P3HT) may be well described without the use of a polarizable environment.

## 4.3 Energetic Disorder

In OSC devices, the charge transport normally can be well described by a hopping mechanism. The charge hopping rates are mostly affected by alignment of electronic transport levels of individual sites (site energies) (Coropceanu et al. 2007). As a result of amorphous structures of OSC materials, there is an energetic disorder in site energies. As discussed in Sect. 3.2, there are two different sources of energetic disorder, static and dynamic energetic disorder, and they affect the charge transport properties in different ways. For example, as pointed out by Tummala et al. (2015), static disorder is always detrimental for charge transport, while dynamic disorder can sometimes lead to an enhancement of the transport properties via a

phonon-assisted contribution to charge-carrier mobility. Therefore, it is important to separately estimate the effects of static and dynamic disorder on charge transport in OSCs.

However, widely used disorder models like the Gaussian disorder model (Bässler 1993) and its extensions (Novikov et al. 1998; Bouhassoune et al. 2009; Kordt and Andrienko 2016) do not make an explicit distinction between these two types of disorder. Meanwhile, many charge transport simulation studies use parametrizations that assume all the disorder to be static (Rühle et al. 2011; Friederich et al. 2015; Yavuz et al. 2015). Only very recently, there are studies simulating static and dynamic energetic disorder in OSC devices. Tummala et al. (2015) investigated the energetic disorder in fullerene systems, including $C_{60}$, $C_{70}$, $PC_{61}BM$, and $PC_{71}BM$, using combined MD and DFT calculations. They found in the $C_{60}$ and $C_{70}$ bulk amorphous structures, the extent of static disorder is comparable to that of dynamic disorder. However, in the case of $PC_{61}BM$ and $PC_{71}BM$, the static disorder increases significantly and is twice as large as the dynamic disorder, due to the conformational changes from the phenyl-butyric acid methyl ester adducts.

de Silva and Van Voorhis (2018) performed combined MD and polarizable QM/MM simulations to study energetic disorder in emission layers of two OLED devices, where PIC-TRZ2 (guest) is dispersed into two different host matrices: mCP and UGH2. From sampling 20 different guest and host sites and their time evolution over 100 MD snapshots, the authors obtained the overall energetic disorder. The choice of the host matrix significantly affects the degree of energetic disorder in site energies for both guest and host components. For example, the standard deviations of EA levels of host and guest in mCP are in the range of 0.15–0.16 eV, while in UGH2 the standard deviations are only 0.10 eV. The authors performed further analysis of ensemble and time distributions of site energies to decompose the total disorder into static and dynamic contributions. Similar to the results of Tummala et al., they showed the dynamic disorder and static disorder are on the same order of magnitude. More interestingly, they found the larger disorder of guest and host molecules in PIC-TRZ2@mCP film is caused by the increase in the static component. The dynamic disorder is not affected by the choice of host, but there is a twofold increase in the static disorder when mCP is used as the host instead of UGH2. The authors thus concluded that polarity is the dominant factor for static disorder, while dynamic disorder does not depend significantly on intermolecular interactions.

## 4.4 Reaction Coordinate

In host-guest OLED devices, guest (emitter) molecules can be treated as impurities because they are dispersed into host molecules at a low concentration. The host-guest interactions have been shown to be important for improving the OLED device performance (Tao et al. 2011; Einzinger et al. 2017). It is then interesting to investigate if the impurities affect the energy levels of their neighboring host molecules, which may further influence the charge transfer efficiency from the host

to the guest. This is similar to band bending at the organic/organic interfaces in OPVs, which can facilitate or inhibit the charge separation processes. As OPV devices normally consist of layers of donor and acceptor molecules, it is easy to define the geometrical distance from a position in the bulk to the interface. Such geometrical distance is then used as the reaction coordinate to study the energy level changes (Yost et al. 2011; Yost and Van Voorhis 2013; Kohn et al. 2017). However, distance between molecules is poorly defined in OLEDs because the molecules are about the same size ($\sim$1 nm) as the space between their centers of mass (1–5 nm). In addition, since many guest molecules exist in the emission layer, it is ambiguous to determine which host-guest distance to use for a specific host molecule. Therefore, a new reaction coordinate is needed to indicate the effective distance between guest molecules and a specific host molecule.

Zhu and Van Voorhis (2016) introduced a new reaction coordinate while studying a phosphorescent OLED host-guest system consisting of BTDF and Ir(ppy)$_3$ using combined MD and polarizable QM/MM simulations. In order to investigate the charge recombination process from the host to the guest, the authors used the "host-guest interaction energy" as the reaction coordinate to evaluate the effect of impurities on the host energy levels. For a specifically chosen host molecule, the host-guest interaction energy is defined as the change of its IP/EA values ($\Delta$IP/$\Delta$EA) with or without the electrostatic interactions with all guest molecules in the system:

$$\Delta IP = |IP\,(on) - IP\,(off)|$$
$$\Delta EA = |EA\,(on) - EA\,(off)| \tag{28}$$

where "off" means point charges and Drude particles on all guest molecules were set to 0 in the MM part of the Hamiltonian in the polarizable QM/MM simulation and "on" means all MM parametrized force field values are kept. Because the electrostatic interaction between two molecules is inversely proportional to their distance, this "host-guest interaction energy" can be used to directly measure the effective distance: the larger host-guest interaction energy means closer distance between the chosen host molecule and all guest molecules.

The IP and EA values of the host as a function of the host-guest interaction energy are shown in Fig. 4. The authors showed that EA of the host increases by about 0.1 eV when the host is closer to guest molecules, while IP of the host stays almost the same. Meanwhile, there are broader distributions of IP/EA at closer host-guest distances, suggesting guest molecules result in slightly more disordered molecular environment. Based on these results, the authors concluded that the change in EA distributions makes the electron transfer from the host to the guest less efficient. This study thus suggests the impurity effect on host energy levels caused by guest molecules is minor in OLEDs. This is also confirmed by a later study of de Silva and Van Voorhis (2018), where they observed no guest impurity effect on host energy levels for PIC-TRZ2@mCP system. The authors explained that such effect is completely suppressed by both static and dynamic energetic disorder in host materials.

**Fig. 4** Left: Distribution of electron affinities of BTDF as a function of the host-guest interaction energy ΔEA. Right: Distribution of ionization potentials of BTDF as a function of the host-guest interaction energy ΔIP. The colored regions indicate the standard deviations. (Zhu and Van Voorhis 2016)
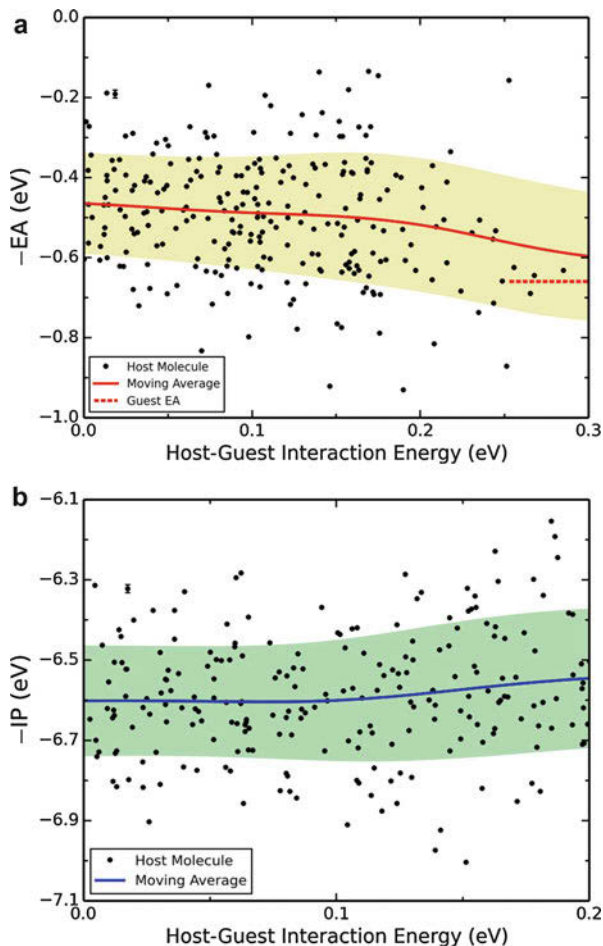


## 4.5    Charge Transfer State

OPV devices need a large driving force to break up excitons and create separate electrons and holes. This driving force is provided by the offset of HOMO/LUMO energy levels of two different OSC materials at their interface. A charge transfer state is formed after the exciton breakup, and the binding energy of the CT state is still around 10 kT, a prohibitively high-energy barrier at room temperature (Zhu et al. 2009). Despite this problem, high interfacial charge separation efficiencies are shown in many OPVs (Kippelen and Brédas 2009; Park et al. 2009; Few et al. 2015). One proposed mechanism is that some of the excess exciton energy might be used to create a "hot" CT state, where the carriers have sufficient kinetic energy to overcome the dissociation barrier. This mechanism is supported by several

theoretical and experimental studies (Dimitrov et al. 2012; Jailaubekov et al. 2012; Bakulin et al. 2013). Tamura and coworkers (Tamura and Burghardt 2013) simulated the exciton dissociation in a TFB/F8BT dimer using a model Hamiltonian and found a significant increase in the free charge formation rate when going through a higher-energy, vibrationally excited CT state. Later work by Shen et al. (2015) using the QM/MM simulation also showed in DTDCTB/fullerene solar cells, the rate of exciton dissociation via the higher-energy CT state is much faster than that of exciton dissociation via the CT ground state.

However, contradictory studies have shown that "cold" CT states can form free charges just as efficiently as "hot" CT states (Vandewal et al. 2013; Bässler and Köhler 2015; Gautam et al. 2016). The experimental study of Lee et al. (2010) on the formation of hot CT states showed that the charge separation efficiency is independent of whether the CT state was formed with excess energy. To understand how charge separates from a cold CT state, it is important to accurately estimate the binding energy of CT states at the organic/organic interface:

$$E_{bind} = \left( E_{IP}^{Donor} - E_{EA}^{Acceptor} \right) - E_{CT} \tag{29}$$

where $E_{CT}$ is the energy of the CT state, $E_{IP}^{Donor}$ and $E_{EA}^{Acceptor}$ are IP of the donor and EA of the acceptor. Time-dependent DFT is known to underestimate CT state energies, while CDFT is shown to be an effective tool to accurately describe CT states and can easily incorporate the environment into the calculations. A restricted open-shell Kohn-Sham approach has also been shown to work well for intramolecular CT states (Hait et al. 2016).

Yost et al. (2011) performed combined MD and polarizable QM/MM simulations on the $H_2Pc$/PTCBI OPV interface. The authors computed the interfacial CT states on a number of nearest-neighbor CT pairs using CDFT and found the CT states have a broader energy distribution than first excited states of either $H_2Pc$ or PTCBI, suggesting the CT states are more sensitive to the electrostatic environment. Through further analysis on the CT binding energy, they found $E_{bind}$ has a clear $R^{-1}$ decay as a function of intermolecular distance, shown in Fig. 5. The average $E_{bind}$ for the closest pairs is 0.2 eV, while averaging over all of the nearest-neighbor pairs gives a $E_{bind}$ of 0.15 eV. There is also a scatter of 0.1 eV within each CT pair due to thermal fluctuations, which is expected to facilitate the initial charge separation at the organic-organic interface. In addition to the CT binding energy, the authors also simulated energy levels at the $H_2Pc$/PTCBI interface, where the band offset is shown to be 0.25 eV larger than the bulk value. The authors thus concluded that the interfacial CT states have higher average energy than fully separated charges in bulk materials but are locally bound at the interface. This study suggested that it is possible for bound interfacial CT states to dissociate in a barrierless fashion without involving "hot" CT states.
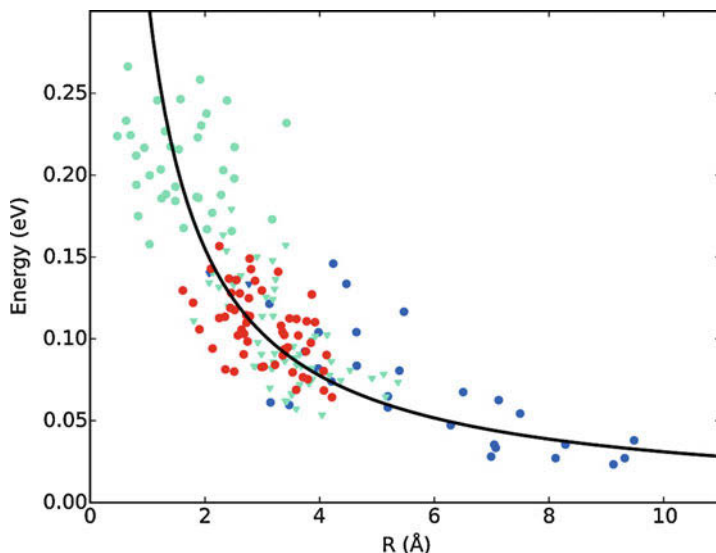
**Fig. 5** Distribution of the $H_2Pc$/PTCBI interfacial CT state binding energies. The coordinate R is a linear combination of intermolecular distances. Each different color/shape combination represents distinct dimer pairs in the simulation cell (Yost et al. 2011)

## 4.6    Charge Mobility

A major factor that limits the efficiency of OPVs is the low charge mobility of OSC materials. The charge mobility at room temperature of most conductive OSC materials is a few tens of $cm^2\ V^{-1}\ s^{-1}$, an order of magnitude smaller than the value of inorganic semiconductors (Ostroverkhova et al. 2006; Troisi 2011). To understand this limit and design more conductive OSCs, one needs to be able to accurately estimate the charge mobility. The simulation of charge mobility in OSCs requires the evaluation of charge transfer rates and a subsequent analysis of the solution of the master equation. Here, we only highlight the computational efforts on charge transfer rates. For the use of kinetic Monte Carlo simulation and coarse-grained lattice models to obtain the charge mobility, we direct the reader to the review by Andrienko and coworkers (Kordt et al. 2015).

To compute the charge transfer rates following Marcus theory, one needs to simulate the band gap, reorganization energy, and electronic coupling of OSC materials, which requires the knowledge of diabatic states and the Hamiltonian of the involved dimer. Semiempirical methods such as ZINDO have also been shown to be useful in computing electronic coupling in a faster manner compared to first principle calculations (Kirkpatrick 2008; Vehoff et al. 2010). However, a more accurate and general method to construct diabatic states is CDFT. Difley et al. (2010) performed CDFT calculations in polarizable MM environment to simulate the charge transfer rates in $Alq_3$ crystalline structure. The authors found the electronic coupling can vary by as much as a factor of 5 as a function of time,

indicating significant dynamic heterogeneity of the hopping rate on the nanosecond time scale. The electronic coupling was also shown to depend strongly on the relative orientation and distance of the involved $Alq_3$ dimer. The authors further found that although the reorganization energy for the electron transfer is larger than the hole transfer, the considerably larger electronic coupling of the electron transfer is enough to overcome the difference in activation barriers. As a result, the electron transfer rate is shown to be larger than the hole transfer, which qualitatively agrees with the experimental observation that $Alq_3$ is a good electron transport material. However, the authors also pointed out one must apply mesoscopic simulations to obtain quantitative predictions on charge mobility (Kwiatkowski et al. 2008) because of the heterogeneity of the charge transfer rates. This study thus suggests it is crucial to include dynamic and static disorder and construct diabatic states accurately when computing the charge transfer rates.

## 5    Conclusion

We have here presented a brief review of the principles of charge transfer in molecular materials. The essential principles of electron transfer can be understood using arguments from Marcus theory – an approximate model that was originally designed for solution-phase electrochemistry but which has seen a shocking degree of success describing charge transfer in widely disparate areas. Next, we highlighted the main ingredients in any computational model: (1) a method to compute localized charge states and their couplings (here chosen to be CDFT), (2) a method to account for environment effects (here chosen to be QM/MM), and (3) a means of differentiating between static and dynamic disorder of the system. Then, using organic LED and PV materials as illustrations, we showed how careful consideration of charge transfer reveals a wealth of information about the system: the behavior of band gaps near heterojunction interfaces; the delocalization length of charge carriers; the effective binding energy of electrons and holes; and the charge mobility of materials.

Moving forward, there are a number of exciting directions for exploration. How do oppositely charged carriers in host-guest complexes interact with one another, and how does that interaction mediate charge recombination? In polymeric systems, how does the dynamic localization and delocalization of charge impact charge transport? How do charges interact with bound electron-hole pairs, like excitons, and how can we understand the rate of exciton-charge annihilation? And finally, how can we combine the molecular and solid-state pictures of charge transport in order to appropriately describe *hybrid* organic-inorganic devices involving molecules and traditional semiconductors simultaneously? All of these are intriguing research directions, and we are at the auspicious moment of history where such questions can and will be answered in the coming years.

# References

Aguilar MA (2001) Separation of the electric polarization into fast and slow components: a comparison of two partition schemes. J Phys Chem A 105:10393–10396. https://doi.org/10.1021/jp011598f

Anderson PW (1958) Absence of diffusion in certain random lattices. Phys Rev 109:1492–1505. https://doi.org/10.1103/PhysRev.109.1492

Appel AM, Bercaw JE, Bocarsly AB et al (2013) Frontiers, opportunities, and challenges in biochemical and chemical catalysis of $CO_2$ fixation. Chem Rev 113:6621–6658. https://doi.org/10.1021/cr300463y

Bakulin AA, Dimitrov SD, Rao A et al (2013) Charge-transfer state dynamics following hole and electron transfer in organic photovoltaic devices. J Phys Chem Lett 4:209–215. https://doi.org/10.1021/jz301883y

Barbara PF, Meyer TJ, Ratner MA (1996) Contemporary issues in electron transfer research. J Phys Chem 100:13148–13168. https://doi.org/10.1021/JP9605663

Barford W, Lidzey DG, Makhov DV, Meijer AJH (2010) Exciton localization in disordered poly(3-hexylthiophene). J Chem Phys 133:44504. https://doi.org/10.1063/1.3459099

Bässler H (1993) Charge transport in disordered organic photoconductors a Monte Carlo simulation study. Phys Status Solidi B 175:15–56. https://doi.org/10.1002/pssb.2221750102

Bässler H, Köhler A (2015) "Hot or cold": how do charge transfer states at the donor–acceptor interface of an organic solar cell dissociate? Phys Chem Chem Phys 17:28451–28462. https://doi.org/10.1039/C5CP04110D

Becke AD (1988) A multicenter numerical integration scheme for polyatomic molecules. J Chem Phys 88:2547–2553. https://doi.org/10.1063/1.454033

Blumberger J, Sprik M (2006) Quantum versus classical electron transfer energy as reaction coordinate for the aqueous $Ru^{2+}/Ru^{3+}$ redox reaction. Theor Chem Accounts 115:113–126. https://doi.org/10.1007/s00214-005-0058-0

Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for {M}onte {C}arlo simulation of {I}sing spin systems. J Comput Phys 17:10–18. https://doi.org/10.1016/0021-9991(75)90060-1

Bouhassoune M, van Mensfoort SLM, Bobbert PA, Coehoorn R (2009) Carrier-density and field-dependent charge-carrier mobility in organic semiconductors with correlated Gaussian disorder. Org Electron 10:437–445. https://doi.org/10.1016/J.ORGEL.2009.01.005

Brédas J-L, Beljonne D, Veaceslav Coropceanu A, Cornil J (2004) Charge-transfer and energy-transfer processes in π-conjugated oligomers and polymers: a molecular picture. Chem Rev 104:4971–5003. https://doi.org/10.1021/CR040084K

Castet F, D'Avino G, Muccioli L et al (2014) Charge separation energetics at organic heterojunctions: on the role of structural and electrostatic disorder. Phys Chem Chem Phys 16:20279–20290. https://doi.org/10.1039/C4CP01872A

Cheng C, Geng H, Yi Y, Shuai Z (2017) Super-exchange-induced high performance charge transport in donor–acceptor copolymers. J Mater Chem C 5:3247–3253. https://doi.org/10.1039/C6TC05534F

Coropceanu V, Cornil J, Filho DA d S et al (2007) Charge transport in organic semiconductors. Chem Rev 107:926–9652. https://doi.org/10.1021/CR050140X

Cortona P (1991) Self-consistently determined properties of solids without band-structure calculations. Phys Rev B 44:8454–8458. https://doi.org/10.1103/PhysRevB.44.8454

Cramer CJ, Truhlar DG (1999) Implicit solvation models: equilibria, structure, spectra, and dynamics. Chem Rev 99:2161–2200. https://doi.org/10.1021/cr960149m

de Silva P, Van Voorhis T (2018) QM/MM study of static and dynamic energetic disorder in the emission layer of an organic light-emitting diode. J Phys Chem Lett 1329–1334. https://doi.org/10.1021/acs.jpclett.8b00040

Difley S, Wang L-P, Yeganeh S et al (2010) Electronic properties of disordered organic semiconductors via QM/MM simulations. Acc Chem Res 43:995–1004. https://doi.org/10.1021/ar900246s

Dimitrov SD, Bakulin AA, Nielsen CB et al (2012) On the energetic dependence of charge separation in low-band-gap polymer/fullerene blends. J Am Chem Soc 134:18189–18192. https://doi.org/10.1021/ja308177d

Einzinger M, Zhu T, de Silva P et al (2017) Shorter exciton lifetimes via an external heavy-atom effect: alleviating the effects of bimolecular processes in organic light-emitting diodes. Adv Mater 29:1701987. https://doi.org/10.1002/adma.201701987

Few S, Frost JM, Nelson J (2015) Models of charge pair generation in organic solar cells. Phys Chem Chem Phys 17:2311–2325. https://doi.org/10.1039/C4CP03663H

Field MJ, Bash PA, Karplus M (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem 11:700–733. https://doi.org/10.1002/jcc.540110605

Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications, 2nd edn. Academic, San Diego

Friederich P, Meded V, Symalla F et al (2015) QM/QM approach to model energy disorder in amorphous organic semiconductors. J Chem Theory Comput 11:560–567. https://doi.org/10.1021/ct501023n

Gautam BR, Younts R, Li W et al (2016) Charge photogeneration in organic photovoltaics: role of hot versus cold charge-transfer excitons. Adv Energy Mater 6:1301032. https://doi.org/10.1002/aenm.201501032

Goldey MB, Brawand NP, Vörös M, Galli G (2017) Charge transport in nanostructured materials: implementation and verification of constrained density functional theory. J Chem Theory Comput 13:2581–2590. https://doi.org/10.1021/acs.jctc.7b00088

Goodenough JB, Park K-S (2013) The Li-ion rechargeable battery: a perspective. J Am Chem Soc 135:1167–1176. https://doi.org/10.1021/ja3091438

Hait D, Zhu T, McMahon DP, Van Voorhis T (2016) Prediction of excited-state energies and singlet–triplet gaps of charge-transfer states using a restricted open-Shell Kohn–Sham approach. J Chem Theory Comput 12:3353–3359. https://doi.org/10.1021/acs.jctc.6b00426

Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theor Chim Acta 44:129–138. https://doi.org/10.1007/BF00549096

Hüfner S (2003) Photoelectron spectroscopy: principles and applications. Springer, Berlin

Hush NS (1961) Adiabatic theory of outer sphere electron-transfer reactions in solution. Trans Faraday Soc 57:557. https://doi.org/10.1039/tf9615700557

Idé J, Mothy S, Savoyant A et al (2013) Interfacial dipole and band bending in model pentacene/C$_{60}$ heterojunctions. Int J Quantum Chem 113:580–584. https://doi.org/10.1002/qua.24006

Ishii H, Sugiyama K, Ito E, Seki K (1999) Energy level alignment and interfacial electronic structures at organic/metal and organic/organic interfaces. Adv Mater 11:605–625. https://doi.org/10.1002/(SICI)1521-4095(199906)11:8<605::AID-ADMA605>3.0.CO;2-Q

Jacob CR, Neugebauer J (2014) Subsystem density-functional theory. Wiley Interdiscip Rev Comput Mol Sci 4:325–362. https://doi.org/10.1002/wcms.1175

Jailaubekov AE, Willard AP, Tritsch JR et al (2012) Hot charge-transfer excitons set the time limit for charge separation at donor/acceptor interfaces in organic photovoltaics. Nat Mater 12:66–73. https://doi.org/10.1038/nmat3500

Jamieson FC, Domingo EB, McCarthy-Ward T et al (2012) Fullerene crystallisation as a key driver of charge separation in polymer/fullerene bulk heterojunction solar cells. Chem Sci 3:485–492. https://doi.org/10.1039/C1SC00674F

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236. https://doi.org/10.1021/JA9621760

Jortner J, Bixon M (1988) Intramolecular vibrational excitations accompanying solvent-controlled electron transfer reactions. J Chem Phys 88:167–170. https://doi.org/10.1063/1.454632

Kaduk B, Kowalczyk T, Van Voorhis T (2012) Constrained density functional theory. Chem Rev 112:321–370. https://doi.org/10.1021/cr200148b

Kelly CP, Cramer CJ, Truhlar DG (2005) SM6: a density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. J Chem Theory Comput 1:1133–1152. https://doi.org/10.1021/ct050164b

Kippelen B, Brédas J-L (2009) Organic photovoltaics. Energy Environ Sci 2:251. https://doi.org/10.1039/b812502n

Kirkpatrick J (2008) An approximate method for calculating transfer integrals based on the ZINDO Hamiltonian. Int J Quantum Chem 108:51–56. https://doi.org/10.1002/qua.21378

Kittel C (2005) Introduction to solid state physics, 8th edn. Wiley, Hoboken

Klamt A (2011) The COSMO and COSMO-RS solvation models. Wiley Interdiscip Rev Comput Mol Sci 1:699–709. https://doi.org/10.1002/wcms.56

Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2:799–805. https://doi.org/10.1039/P29930000799

Kohn AW, McMahon DP, Wen S, Van Voorhis T (2017) The impact of carrier delocalization and interfacial electric field fluctuations on organic photovoltaics. J Phys Chem C 121:26629–26636. https://doi.org/10.1021/acs.jpcc.7b08726

Kojima A, Teshima K, Shirai Y, Miyasaka T (2009) Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. J Am Chem Soc 131:6050–6051. https://doi.org/10.1021/ja809598r

Kordt P, Andrienko D (2016) Modeling of spatially correlated energetic disorder in organic semiconductors. J Chem Theory Comput 12:36–40. https://doi.org/10.1021/acs.jctc.5b00764

Kordt P, van der Holst JJM, Al Helwi M et al (2015) Modeling of organic light emitting diodes: from molecular to device properties. Adv Funct Mater 25:1955–1971. https://doi.org/10.1002/adfm.201403004

Kwiatkowski JJ, Nelson J, Li H et al (2008) Simulating charge transport in tris (8-hydroxyquinoline) aluminium (Alq3). Phys Chem Chem Phys 10:1852. https://doi.org/10.1039/b719592c

Lamoureux G, Roux B (2003) Modeling induced polarization with classical Drude oscillators: theory and molecular dynamics simulation algorithm. J Chem Phys 119:3025–3039. https://doi.org/10.1063/1.1589749

Lee J, Vandewal K, Yost SR et al (2010) Charge transfer state versus hot exciton dissociation in polymer−fullerene blended solar cells. J Am Chem Soc 132:11878–11880. https://doi.org/10.1021/ja1045742

Lin H, Truhlar DG (2007) QM/MM: what have we learned, where are we, and where do we go from here? Theor Chem Accounts 117:185–199. https://doi.org/10.1007/s00214-006-0143-z

Linares M, Beljonne D, Cornil J et al (2010) On the interface dipole at the Pentacene−fullerene heterojunction: a theoretical study. J Phys Chem C 114:3215–3224. https://doi.org/10.1021/jp910005g

Marcus RA (1956) On the theory of oxidation-reduction reactions involving electron transfer. I. J Chem Phys 24:966–978. https://doi.org/10.1063/1.1742723

Marcus RA (1964) Chemical and electrochemical electron-transfer theory. Annu Rev Phys Chem 15:155–196. https://doi.org/10.1146/annurev.pc.15.100164.001103

Marcus RA (1993) Electron transfer reactions in chemistry. Theory and experiment. Rev Mod Phys 65:599–610. https://doi.org/10.1103/RevModPhys.65.599

Marcus RA, Sutin N (1985) Electron transfers in chemistry and biology. Biochim Biophys Acta Rev Bioenerg 811:265–322. https://doi.org/10.1016/0304-4173(85)90014-X

Marenich AV, Olson RM, Kelly CP et al (2007) Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. J Chem Theory Comput 3:2011–2033. https://doi.org/10.1021/ct7001418

Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113:6378–6396. https://doi.org/10.1021/jp810292n

McMahon DP, Troisi A (2010) Organic semiconductors: impact of disorder at different timescales. ChemPhysChem 11:2067–2074. https://doi.org/10.1002/cphc.201000182

McMahon DP, Cheung DL, Troisi A (2011) Why holes and electrons separate so well in polymer/fullerene photovoltaic cells. J Phys Chem Lett 2:2737–2741. https://doi.org/10.1021/jz201325g

Mead CA, Truhlar DG (1982) Conditions for the definition of a strictly diabatic electronic basis for molecular systems. J Chem Phys 77:6090–6098. https://doi.org/10.1063/1.443853

Mennucci B (2012) Polarizable continuum model. WIREs Comput Mol Sci 2:386–404. https://doi.org/10.1002/wcms.1086

Miertuš S, Scrocco E, Tomasi J (1981) Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem Phys 55:117–129. https://doi.org/10.1016/0301-0104(81)85090-2

Miller A, Abrahams E (1960) Impurity conduction at low concentrations. Phys Rev 120:745–755. https://doi.org/10.1103/PhysRev.120.745

Nakamura S, Mukai T, Senoh M (1994) Candela-class high-brightness InGaN/AlGaN double-heterostructure blue-light-emitting diodes. Appl Phys Lett 64:1687–1689. https://doi.org/10.1063/1.111832

Nan G, Wang L, Yang X et al (2009) Charge transfer rates in organic semiconductors beyond first-order perturbation: from weak to strong coupling regimes. J Chem Phys 130:24704. https://doi.org/10.1063/1.3055519

Nayak PK, Narasimhan KL, Cahen D (2013) Separating charges at organic interfaces: effects of disorder, hot states, and electric field. J Phys Chem Lett 4:1707–1717. https://doi.org/10.1021/jz4002339

Newton MD (1991) Quantum chemical probes of electron-transfer kinetics: the nature of donor-acceptor interactions. Chem Rev 91:767–792. https://doi.org/10.1021/cr00005a007

Novikov SV, Dunlap DH, Kenkre VM et al (1998) Essential role of correlations in governing charge transport in disordered organic materials. Phys Rev Lett 81:4472–4475. https://doi.org/10.1103/PhysRevLett.81.4472

Ostroverkhova O, Cooke DG, Hegmann FA et al (2006) Ultrafast carrier dynamics in pentacene, functionalized pentacene, tetracene, and rubrene single crystals. Appl Phys Lett 88:162101. https://doi.org/10.1063/1.2193801

Park SH, Roy A, Beaupré S et al (2009) Bulk heterojunction solar cells with internal quantum efficiency approaching 100%. Nat Photonics 3:297–302. https://doi.org/10.1038/nphoton.2009.69

Pasveer WF, Cottaar J, Tanase C et al (2005) Unified description of charge-carrier mobilities in disordered semiconducting polymers. Phys Rev Lett 94:206601. https://doi.org/10.1103/PhysRevLett.94.206601

Pavanello M, Neugebauer J (2011) Modelling charge transfer reactions with the frozen density embedding formalism. J Chem Phys 135:234103. https://doi.org/10.1063/1.3666005

Pavanello M, Van Voorhis T, Visscher L, Neugebauer J (2013) An accurate and linear-scaling method for calculating charge-transfer excitation energies and diabatic couplings. J Chem Phys 138:054101. https://doi.org/10.1063/1.4789418

Rühle V, Lukyanov A, May F et al (2011) Microscopic simulations of charge transport in disordered organic semiconductors. J Chem Theory Comput 7:3335–3345. https://doi.org/10.1021/ct200388s

Sariciftci NS, Smilowitz L, Heeger AJ et al (1992) Photoinduced electron transfer from a conducting polymer to buckminsterfullerene. Science 258:1474–1476. https://doi.org/10.1126/science.258.5087.1474

Savoie BM, Rao A, Bakulin AA et al (2014) Unequal partnership: asymmetric roles of polymeric donor and fullerene acceptor in generating free charge. J Am Chem Soc 136:2876–2884. https://doi.org/10.1021/ja411859m

Schober C, Reuter K, Oberhofer H (2016) Critical analysis of fragment-orbital DFT schemes for the calculation of electronic coupling values. J Chem Phys 144:54103. https://doi.org/10.1063/1.4940920

Senthilkumar K, Grozema FC, Bickelhaupt FM, Siebbeles LDA (2003) Charge transport in columnar stacked triphenylenes: effects of conformational fluctuations on charge transfer integrals and site energies. J Chem Phys 119:9809–9817. https://doi.org/10.1063/1.1615476

Shen X, Han G, Fan D et al (2015) Hot charge-transfer states determine exciton dissociation in the DTDCTB/$C_{60}$ complex for organic solar cells: a theoretical insight. J Phys Chem C 119:11320–11326. https://doi.org/10.1021/jp512574d

Shuai Z, Geng H, Xu W et al (2014) From charge transport parameters to charge mobility in organic semiconductors through multiscale simulation. Chem Soc Rev 43:2662. https://doi.org/10.1039/c3cs60319a

Siddarth P, Marcus RA (1993) Electron-transfer reactions in proteins: an artificial intelligence approach to electronic coupling. J Phys Chem 97:2400–2405. https://doi.org/10.1021/j100112a047

Stone AJ (1981) Distributed multipole analysis, or how to describe a molecular charge distribution. Chem Phys Lett 83:233–239. https://doi.org/10.1016/0009-2614(81)85452-8

Tachiya M (1993) Generalization of the Marcus equation for the electron-transfer rate. J Phys Chem 97:5911–5916. https://doi.org/10.1021/j100124a023

Tamura H, Burghardt I (2013) Ultrafast charge separation in organic photovoltaics enhanced by charge delocalization and vibronically hot exciton dissociation. J Am Chem Soc 135:16364–16367. https://doi.org/10.1021/ja4093874

Tang CW, VanSlyke SA (1987) Organic electroluminescent diodes. Appl Phys Lett 51:913–915. https://doi.org/10.1063/1.98799

Tao Y, Yang C, Qin J (2011) Organic host materials for phosphorescent organic light-emitting diodes. Chem Soc Rev 40:2943–2970. https://doi.org/10.1039/c0cs00160k

Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105:2999–3093

Troisi A (2011) Charge transport in high mobility molecular semiconductors: classical models and new theories. Chem Soc Rev 40:2347. https://doi.org/10.1039/c0cs00198h

Troisi A, Orlandi G (2006a) Charge-transport regime of crystalline organic semiconductors: diffusion limited by thermal off-diagonal electronic disorder. Phys Rev Lett 96:86601. https://doi.org/10.1103/PhysRevLett.96.086601

Troisi A, Orlandi G (2006b) Dynamics of the intermolecular transfer integral in crystalline organic semiconductors. J Phys Chem A 110:4065–4070. https://doi.org/10.1021/JP055432G

Tummala NR, Zheng Z, Aziz SG et al (2015) Static and dynamic energetic disorders in the $C_{60}$, $PC_{61}$ BM, $C_{70}$, and $PC_{71}$ BM fullerenes. J Phys Chem Lett 6:3657–3662. https://doi.org/10.1021/acs.jpclett.5b01709

Van Voorhis T, Kowalczyk T, Kaduk B et al (2010) The diabatic picture of electron transfer, reaction barriers, and molecular dynamics. Annu Rev Phys Chem 61:149–170. https://doi.org/10.1146/annurev.physchem.012809.103324

Vandewal K, Albrecht S, Hoke ET et al (2013) Efficient charge generation by relaxed charge-transfer states at organic interfaces. Nat Mater 13:63–68. https://doi.org/10.1038/nmat3807

Vehoff T, Baumeier B, Troisi A, Andrienko D (2010) Charge transport in organic crystals: role of disorder and topological connectivity. J Am Chem Soc 132:11702–11708. https://doi.org/10.1021/ja104380c

Verlaak S, Beljonne D, Cheyns D et al (2009) Electronic structure and geminate pair energetics at organic-organic interfaces: the case of Pentacene/C60 heterojunctions. Adv Funct Mater 19:3809–3814. https://doi.org/10.1002/adfm.200901233

Wang Y, Chen KS, Mishler J et al (2011) A review of polymer electrolyte membrane fuel cells: technology, applications, and needs on fundamental research. Appl Energy 88:981–1007. https://doi.org/10.1016/J.APENERGY.2010.09.030

Warshel A (1982) Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions. J Phys Chem 86:2218–2224. https://doi.org/10.1021/j100209a016

Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103:227–249. https://doi.org/10.1016/0022-2836(76)90311-9

Wesolowski TA, Warshel A (1993) Frozen density functional approach for ab initio calculations of solvated molecules. J Phys Chem 97:8050–8053. https://doi.org/10.1021/j100132a040

Wesolowski TA, Weber J (1996) Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules. Chem Phys Lett 248:71–76. https://doi.org/10.1016/0009-2614(95)01281-8

Wu Q, Van Voorhis T (2005) Direct optimization method to study constrained systems within density-functional theory. Phys Rev A 72:24502. https://doi.org/10.1103/PhysRevA.72.024502

Wu Q, Van Voorhis T (2006) Extracting electron transfer coupling elements from constrained density functional theory. J Chem Phys 125:164105. https://doi.org/10.1063/1.2360263

Yavuz I, Martin BN, Park J, Houk KN (2015) Theoretical study of the molecular ordering, paracrystallinity, and charge mobilities of oligomers in different crystalline phases. J Am Chem Soc 137:2856–2866. https://doi.org/10.1021/ja5076376

Yeganeh S, Ratner MA (2006) Effects of anharmonicity on nonadiabatic electron transfer: a model. J Chem Phys 124:44108. https://doi.org/10.1063/1.2162172

Yost SR, Van Voorhis T (2013) Electrostatic effects at organic semiconductor interfaces: a mechanism for "cold" exciton breakup. J Phys Chem C 117:5617–5625. https://doi.org/10.1021/jp3125186

Yost SR, Wang L-P, Van Voorhis T (2011) Molecular insight into the energy levels at the organic donor/acceptor interface: a quantum mechanics/molecular mechanics study. J Phys Chem C 115:14431–14436. https://doi.org/10.1021/jp203387m

Yost SR, Hontz E, McMahon DP, Van Voorhis T (2014) Electronic and optical properties at organic/organic interfaces in organic solar cells. In: Multiscale modelling of organic and hybrid photovoltaics. Topics in current chemistry, Beljonne D, Cornil J (eds), vol 352. Springer, Berlin Heidelberg, pp 103–150

Zahn DRT, Gavrila GN, Gorgoi M (2006) The transport gap of organic semiconductors studied using the combination of direct and inverse photoemission. Chem Phys 325:99–112. https://doi.org/10.1016/J.CHEMPHYS.2006.02.003

Zhu T, Van Voorhis T (2016) Charge recombination in phosphorescent organic light-emitting diode host–guest systems through QM/MM simulations. J Phys Chem C 120:19987–19994. https://doi.org/10.1021/acs.jpcc.6b05559

Zhu X-Y, Yang Q, Muntwiler M (2009) Charge-transfer excitons at organic semiconductor surfaces and interfaces. Acc Chem Res 42:1779–1787. https://doi.org/10.1021/ar800269u

# Van der Waals Interactions in Material Modelling

**13**

Jan Hermann and Alexandre Tkatchenko

## Contents

**Abstract**

Van der Waals (vdW) interactions stem from electronic zero-point fluctuations and are often critical for the correct description of structure, stability, and response properties of molecules and materials, including biomolecules, nanomaterials, and material interfaces. Here, we give a conceptual as well as mathematical overview of the current state of modeling vdW interactions,

J. Hermann · A. Tkatchenko (✉)
Physics and Materials Science Research Unit, University of Luxembourg, Luxembourg, Luxembourg
e-mail: alexandre.tkatchenko@uni.lu

259

focusing in particular on the consequences of different approximations for practical applications. We present a systematic classification of approximate first-principles models based on the adiabatic-connection fluctuation-dissipation theorem, namely the nonlocal density functionals, interatomic methods, and methods based on the random-phase approximation. The applicability of these methods to different types of materials and material properties is discussed in connection with availability of theoretical and experimental benchmarks. We conclude with a roadmap of the open problems that remain to be solved to construct a universal, efficient, and accurate vdW model for realistic material modeling.

# 1 Introduction

van der Waals (vdW) interactions stem from electronic zero-point (and possibly thermal) fluctuations in electronic matter (Langbein 1974; Parsegian 2005). Therefore, vdW interactions scale rapidly with system size and are often critical for the correct description of structure, stability, and response properties of molecules and materials, including biomolecules, nanomaterials, and material interfaces. This makes proper description of vdW interactions a crucial aspect of modern material modelling. This is especially important in the context of electronic-structure calculations using approximate density functionals. Such functionals are normally semilocal ("short sighted") in the electronic density, meaning that nonlocal vdW interactions are poorly or not at all described by these functionals.

From a classical electrostatic perspective, electrons repel each other. However, quantum-mechanical correlation effects typically act to minimize the electric repulsion between electrons. This *correlation* in the electronic motion results in instantaneous effective multipoles that interact via electrostatic forces, attracting different regions toward each other. When there is no significant overlap between two interacting regions of electronic matter, we usually speak of the long-range electron correlation, which is the underlying microscopic cause of vdW interactions. (As Margenau put it already in 1939, "the term 'van der Waals force' is not one of very precise usage," which holds to this date. Here, we will use the term exclusively to refer to the electron correlation part of noncovalent interactions.)

In many approaches to material modelling, ranging from empirical to those based on first principles, the models that describe the short-range and long-range parts of the electron correlation are constructed separately, because each requires a different set of considerations. This chapter discusses currently used approaches to the calculation of the long-range correlation energy – the vdW energy – with special focus on the case when the short-range part of the correlation energy (and other total energy components) is calculated within the density functional theory (DFT). We begin with a conceptual understanding and importance of vdW interactions in materials (Sect. 2), which has motivated the recent rapid development of new vdW models presented in the subsequent sections. This is followed by a brief conceptual overview of existing and widely used vdW models (Sect. 3), which can serve

either as a stand-alone short synopsis or as an introduction to the subsequent more detailed exposition. We then continue with a more theoretical presentation of current approaches to modelling vdW interactions, first introducing a general framework for discussing vdW models (Sect. 4), followed by a classification of current approaches within that framework (Sect. 5). None of the existing vdW models is the best choice for every system, and a close attention must be paid to their accuracy and areas of applicability (Sect. 6). We conclude with an outline of outstanding problems in the topic of modelling vdW interactions from first principles of quantum mechanics (Sect. 7).

## 2    Role of van der Waals Interactions in Materials

Functional materials are becoming increasingly smaller in size and more heterogeneous in composition. These two aspects of novel nanomaterials lead to the emergence of nontrivial quantum-mechanical effects that depend on size and topology and which may ultimately determine the properties of a material of interest. One important consequence of this evolution beyond traditional materials, the functionality of which was largely regulated by bulk observables, is that noncovalent interactions play an increasingly important role in determining the structure, stability, and ensuing function of homogeneous and heterogeneous nanostructured materials. van der Waals interactions, which exhibit nontrivial scaling behavior with system size (Dobson et al. 2006; Ruzsinszky et al. 2012; Gobre and Tkatchenko 2013), are often the dominant part of such noncovalent interactions. In general, vdW interactions have already been recognized as playing an instrumental role in determining the structure, stability, and functionality of biological materials, supramolecular and sensor chemistry, pharmaceuticals, dye-sensitized solar cells, and many other systems. More recently, the field of "van der Waals heterostructures" has moved into the forefront (Geim and Grigorieva 2013) and has already led to fundamental advances in the study of low-dimensional materials and to a number of novel technological applications.

In this context, the importance of understanding and accurately modelling vdW interactions in realistic materials can hardly be overemphasized. However, our ability to model these ubiquitous quantum mechanical effects has been severely impeded by the prohibitively high computational cost of explicitly correlated quantum chemical methods and the lack of efficient approximations to the many-electron correlation problem for large systems (Szabo and Ostlund 1996). In fact, most successful approximations employed for modelling vdW interactions in materials rely on the rather crude lowest-order pairwise additive approximation, which is only exact in the weak polarization limit and at large interatomic distances. For condensed-phase systems with a moderate to large polarizability density, such pairwise approximations can often lead to qualitatively incorrect predictions of structural, energetic, and response properties.

The structure and binding in organic materials is often driven by vdW interactions, in combination with other contributions such as repulsion, electrostatics,

and induction. Therefore, accurate modelling of vdW interactions is critical for understanding the properties of organic materials. In recent years, substantial progress has been achieved in the theoretical prediction of structures and stabilities of molecular crystals by using vdW-inclusive DFT approaches. Today, the structures of (simple) organic molecular crystals can be predicted with an accuracy of 2–3% and cohesive energies to 1–2 kcal/mol. Proper description of vdW correlations becomes even more relevant when one looks at the relative energetics of molecular systems, which are essential to predict the polymorphic behavior of molecular crystals.

While the crucial role of vdW interactions in organic materials is well established, our understanding of the relative importance of these ubiquitous interactions in semiconductors, ionic solids, and metals is still in development. The contribution of the long-range vdW energy to the cohesive energy of elemental and binary semiconductors and ionic solids amounts to 0.2–0.3 eV/atom, which is around 8% of the cohesive energy (Zhang et al. 2011). The contribution of vdW energy to the bulk modulus is even more pronounced, reaching up to 22% for Ge and GaAs. Notably, the inclusion of vdW interactions in DFA calculations allows to simultaneously improve the performance for lattice constants, cohesive energies, and bulk moduli, when compared to experiment. Similar conclusions have been reached for a wide variety of hard solids. Because vdW interactions typically have larger contributions to relative energetics than absolute ones, one expects significant effects for phase transition pressures and phase diagrams of most solids.

The properties of many materials are substantially affected by the presence of simple and complex defects. For example, the properties of semiconductors are largely determined by neutral and charged interstitials and vacancies (Freysoldt et al. 2014). The formation of defects entails a modification of polarization around defect sites, and this can have a substantial effect on the contribution of vdW energy to the stability and mobility of defects. For instance, the inclusion of interactions in DFT improves the description of defect formation energies, significantly changes the barrier geometries for defect diffusion, and brings migration barrier heights into close agreement with experimental values (Gao and Tkatchenko 2013). In the case of Si, the vdW energy substantially decreases the migration barriers of interstitials and impurities by up to 0.4 eV, qualitatively changing the diffusion mechanism. Moving beyond point defects, it is to be expected that more complex neutral and charged multiatom defects and dislocations will lead to even stronger nonlocal polarization effects and intricate dependence of vdW interactions on the nature of defects.

Hybrid inorganic/organic systems (HIOS) are relevant for many applications in catalysis, light-emitting diodes, single-molecule junctions, molecular sensors and switches, and photovoltaics. The predictive modelling and understanding of the structure and stability of such hybrid systems are an essential prerequisite for tuning their electronic properties and functions. The bonding in HIOS is often determined by a delicate balance between covalent bonds, hydrogen bonds, charge transfer, Pauli repulsion, and vdW interactions. Yet the latter are at the edge of what is possible with current vdW approaches, because the inorganic substrate is often a

metal or a doped semiconductor, which is difficult to model due to the delocalized nature of metallic single-particle excitations.

## 3 Overview of Existing van der Waals Models

A common starting point for virtually all existing vdW models is an exact expression for the electron correlation energy that can be constructed from the ACFD theorem (Gunnarsson and Lundqvist 1976; Langreth and Perdew 1977). The essential idea of this expression is an interpolation between a reference noninteracting mean-field system and the fully interacting many-body system. The starting quantity from which the correlation energy can be calculated is some form of a response function (nonlocal polarizability, dielectric function, density response function), which describes the linear response of the system to external electric field. This quantity can be calculated either from first principles, using some mean-field electronic-structure theory such as the Hartree–Fock (HF) method or Kohn–Sham DFT (KS-DFT), or constructed semiempirically using approximate models. The former approach is less ambiguous, but the complexity of such ab initio response functions demands large computational cost and even then requires substantial approximations to the interelectronic interactions. In contrast, the latter approach leads to efficient methods, in which much of the complexity of the electronic structure can be treated effectively by suitable parameterization of the model response functions. In both cases, either the response functions or the interelectronic interactions (or both) are treated approximately, leading to deviations from the in-principle exact ACFD correlation energy.

Essentially all existing methods for modelling vdW interactions can be derived by following various approximations to the ACFD theorem. Two major classes of approximations that can be identified in existing vdW models are the coarse-graining of the response functions and truncation of the many-body expansion. In their full form, the response functions are two-point spatial functions, specifying a response of a system at some point to the perturbation in the electric field in another point. Such a description enables the calculation of the electron correlations across the whole range of interelectronic distances, including the short-range intra-atomic as well as long-range intermolecular correlations. This level of detail is largely unnecessary for the calculation of the vdW energy, which comprises only the long-range part of the electron correlation energy. Therefore, a common approach is to coarse-grain the response functions into finite-size fragments – typically atoms – and to evaluate only the long-range correlations between the fragments, while the short-range intrafragment interactions are captured in the effective response properties of the fragments. The second common approximation involves the truncation of the many-body expansion. The closed-form ACFD expression can be expanded as an infinite sum of terms that involve repeated couplings of the noninteracting response functions. Ordering these terms by the number of times a response function occurs in them, the electron correlation energy starts at second order (two occurrences) and continues to infinity. The physical interpretation of the

$n$-th order is such that $n$ electronic fluctuations interact via the long-range Coulomb potential to yield a contribution to the correlation energy. The magnitude of the contribution of the individual orders to the vdW energy decreases with the growing order, which motivates the common approximation of truncating the infinite sum at some order, usually second or third. Using only the lowest second order results in the large class of pairwise methods, which neglect any many-body effects in vdW interactions.
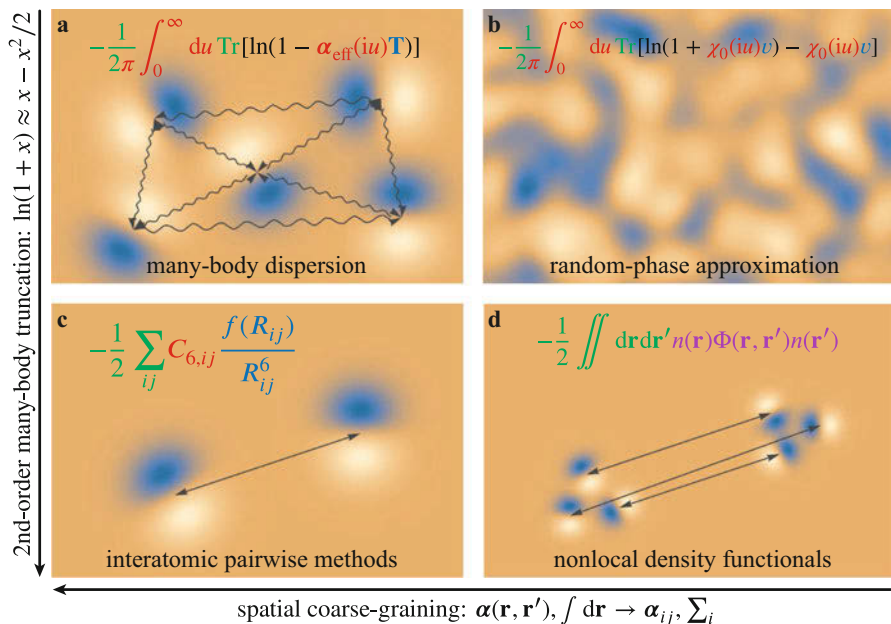


**Fig. 1** Classification of different approaches to modelling vdW interactions based on their approximations to the adiabatic-connection fluctuation–dissipation (ACFD) formula. The light yellow and blue illustrate density fluctuations, and the arrows denote interactions between them. General formulas for each method class are shown, with colors denoting parts with the same origin in the ACFD formula: summation of all interactions in the system (green), encoding of the response properties (red), and the interaction potentials (blue). In random-phase approximation (RPA), both the noninteracting frequency-dependent density response function, $\chi_0(\mathbf{r}, \mathbf{r}', u)$, and the electronic Coulomb interaction, $v(\mathbf{r}, \mathbf{r}') = |\mathbf{r} - \mathbf{r}'|^{-1}$, are nonlocal spatial quantities, and their multiplication should be interpreted as $[\chi_0 v](\mathbf{r}, \mathbf{r}', u) = \int d\mathbf{r}'' \chi_0(\mathbf{r}, \mathbf{r}'', u) v(\mathbf{r}'', \mathbf{r}')$. The RPA itself neglects the exchange–correlation screening of the Coulomb interaction and is not illustrated in this figure. In many-body dispersion (MBD), the response function and potential are coarse-grained and expressed as the nonlocal polarizability matrix, $\boldsymbol{\alpha}_{ij}(u)$, and dipole matrix, $\mathbf{T}_{ij}$, which are multiplied as ordinary matrices. Nonlocal functionals truncate the many-body expansion (expressed as the logarithm in RPA and MBD) at second order, and the response function and interaction potential are intermingled (violet) in the nonlocal kernel, $\Phi(\mathbf{r}, \mathbf{r}')$. Interatomic pairwise models, coarse-grained and truncated at second order, integrate the frequency-dependent polarizability into the so-called $C_6$ coefficients, while the square of the dipole potential yields the $R^{-6}$ power law

Organizing existing vdW models into a 2-by-2 matrix based on whether they use the coarse-graining and/or the many-body truncation approximations (Fig. 1) provides a concise classification that allows for an efficient discussion of the general behavior of these models. The class of methods that do not coarse-grain nor truncate the many-body expansion are based exclusively on the ab initio noninteracting response functions obtained from the HF or KS-DFT methods. This requires other approximations to the interelectronic interactions in the ACFD formula, which make its evaluation tractable (though still computationally demanding). The most straightforward of these approaches is the so-called random-phase approximation (RPA), in which the complex interaction between electronic fluctuations is replaced with the bare Coulomb potential. In contrast, the other three vdW model classes use model noninteracting response functions (usually polarizability), in which the electronic fluctuations are usually assumed to be localized. In nonlocal density functionals, the response function is modelled as a semilocal polarizability functional of the electron density, which is coupled between all pairs of points in space, resulting in a double spatial integral. The many-body dispersion (MBD) approach models the response of atomic fragments as that of harmonic oscillators, which enables efficient evaluation of the ACFD formula without any truncation of the many-body expansion. Finally, interatomic pairwise models use both the coarse-graining and the truncation to provide a particularly simple formula for the vdW energy, in which the response functions are expressed in the form of the so-called dispersion $C_6$ ($C_8$, $C_{10}, \dots$) coefficients.

Within each of the four classes of vdW models, there are multiple instances that differ in more subtle details. These include the particular parametrization of the model response functions or the mechanisms that separate the short-range and long-range parts of the electron correlation. These differences influence the performance of the individual methods but do not change their general behavior for different types of molecules and materials, which will be discussed in more detail in Sect. 6. These individual models together with a more rigorous mathematical formulation of this overview are presented in the following two sections.

## 4   Long-Range Electron Correlation

The ACFD theorem yields an in-principle exact expression for the electron correlation energy and serves as a basis for the various approximate vdW models discussed in this chapter. The present and the following sections give a brief introduction to the mathematical formulation of this topic, which can be found in greater detail elsewhere (Hermann et al. 2017). The ACFD formula expresses the vdW energy in terms of the response of an electronic system to an external electric field, which is introduced in the remainder of this section, while the formula itself and the various approximations to it are discussed in the following section.

## 4.1    Response Functions and Polarizability

The polarization of electronic matter under the influence of a time-periodic external electric field, $\mathbf{E}_\Delta = -\nabla v_\Delta$, with a given frequency, $u$, can be expressed by the change in the electron density, $\Delta n$, from the unpolarized state ($\mathbf{E}_\Delta = 0$). In the linear regime, this change is related to the corresponding potential, $v_\Delta$, via the density response function,

$$\Delta n\left(\mathbf{r}, u\right) = \int \mathrm{d}\mathbf{r}' \chi\left(\mathbf{r}, \mathbf{r}', u\right) v_\Delta\left(\mathbf{r}', u\right) \tag{1}$$

Alternatively, the polarization state can be described by the polarization density, $\mathbf{P}$, which can be interpreted as a dipole density and which gives the polarized charge density via divergence,

$$\Delta n\left(\mathbf{r}, u\right) = \nabla \cdot \mathbf{P}\left(\mathbf{r}, u\right) \tag{2}$$

The polarization density is related to the electric field via the (nonlocal) dipole polarizability, $\boldsymbol{\alpha}$, (Hunt 1983),

$$\mathbf{P}\left(\mathbf{r}, u\right) = -\int \mathrm{d}\mathbf{r}' \boldsymbol{\alpha}\left(\mathbf{r}, \mathbf{r}', u\right) \mathbf{E}_\Delta\left(\mathbf{r}', u\right) \tag{3}$$

In general, the response of the electron density is anisotropic, $\mathbf{E}_\Delta$ and $\mathbf{P}$ are not aligned, and the polarizability must be a tensor. The relation between the density response function and dipole polarizability is obtained by taking the divergence of Eq. 3, integrating by parts, using the definitions of $\mathbf{E}_\Delta$ and $\mathbf{P}$, and comparing to Eq. 1,

$$\begin{aligned}
\chi\left(\mathbf{r}, \mathbf{r}', u\right) &= -\nabla \cdot \nabla' \cdot \boldsymbol{\alpha}\left(\mathbf{r}, \mathbf{r}', u\right) \\
&= -\sum_{\iota\zeta} \frac{\partial^2}{\partial r_\iota \partial r'_\zeta} \alpha_{\iota\zeta}\left(\mathbf{r}, \mathbf{r}', u\right) \quad (\iota, \zeta = x, y, z)
\end{aligned} \tag{4}$$

Whereas the electron density and the density response functions interact electrically via the Coulomb operator, the polarization density and dipole polarizability interact via the dipole operator,

$$\mathbf{T}\left(\mathbf{R}\right) = \nabla \otimes \nabla' v\left(\left|\mathbf{r} - \mathbf{r}'\right|\right)\Big|_{\substack{\mathbf{r}=\mathbf{R} \\ \mathbf{r}'=\mathbf{0}}} = \frac{-3\mathbf{R} \otimes \mathbf{R} + R^2 \mathbf{I}}{R^5} \tag{5}$$

For instance, the electrostatic Coulomb self-interaction of $\Delta n$, with its corresponding $\mathbf{P}$, can be expressed in two equivalent ways,

$$\begin{aligned}
J\left[\Delta n\right] &= \tfrac{1}{2} \iint \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r} \Delta n\left(\mathbf{r}_1\right) v\left(\left|\mathbf{r}_1 - \mathbf{r}_2\right|\right) \Delta n\left(\mathbf{r}_2\right) \\
&= \tfrac{1}{2} \iint \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \mathbf{P}\left(\mathbf{r}_1\right) \cdot \mathbf{T}\left(\mathbf{r}_1 - \mathbf{r}_2\right) \mathbf{P}\left(\mathbf{r}_2\right)
\end{aligned} \tag{6}$$

The description of the response via the density response function and dipole polarizability is equivalent, and likewise the ACFD formula can be expressed using both. The density response functions are directly accessible from ab initio electronic-structure calculations, whereas the dipole polarizability is better suited for the formulation of approximate models. The reason for this is that the density response function has a complex nodal structure, as it describes depletion of the electron density at some points and its accumulation elsewhere. In contrast, the corresponding polarizability is a smooth rotation-free vector field which encodes that underlying nodal structure implicitly in terms of its local behavior via the divergence operators in Eq. 4. This is true even in the case of a delocalized density response that is characteristic of gapless or near-gapless systems. Therefore, the strength of the response is translated directly into the magnitude of the polarizability, whereas it is translated only indirectly into the magnitude of the gradient of the density response function.

For illustration, consider two one-dimensional (1D) Gaussian charge densities located at $\pm 1$ (as crude model of atoms) and two model density response functions, local and nonlocal (Fig. 2). In one dimension, the dipole polarizability is a scalar and uniquely determined by integrating over the density response function,

$$\alpha^{1D}(x, y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \chi^{1D}(x', y') \tag{7}$$

Even in these trivial models, the density response function changes sign around atoms and has a nontrivial nodal structure, whereas the polarizability is positive everywhere. Furthermore, the delocalized density response translates into a polarizability that is still localized but over a larger region spanning both atoms.

## 4.2   Harmonic Oscillator Model of Polarizability

When the dipole polarizability is localized, as in the examples in the previous section, it can be relatively accurately represented as an effective local polarizability, $\alpha^{\mathrm{eff}}(\mathbf{r}, u)$, formally obtained by integrating over some neighborhood, $M(\mathbf{r})$, around each point, $\mathbf{r}$,

$$\alpha^{\mathrm{eff}}(\mathbf{r}, \mathbf{r}', u) \approx \delta(|\mathbf{r} - \mathbf{r}'|) \int_{M(\mathbf{r})} d\mathbf{r}'' \alpha(\mathbf{r}, \mathbf{r}'', u) \equiv \delta(|\mathbf{r} - \mathbf{r}'|) \alpha^{\mathrm{eff}}(\mathbf{r}, u) \tag{8}$$

A wide variety of approaches to model the spatial dependence of $\alpha^{\mathrm{eff}}$ exist, which will be discussed in the following sections.

The frequency dependence, on the other hand, is very often modelled by that of a charged harmonic oscillator (HO) – a Drude oscillator. Having been Fourier-transformed from the time domain, the frequency-dependent polarizability is a complex-valued function, with the real and imaginary parts encoding the
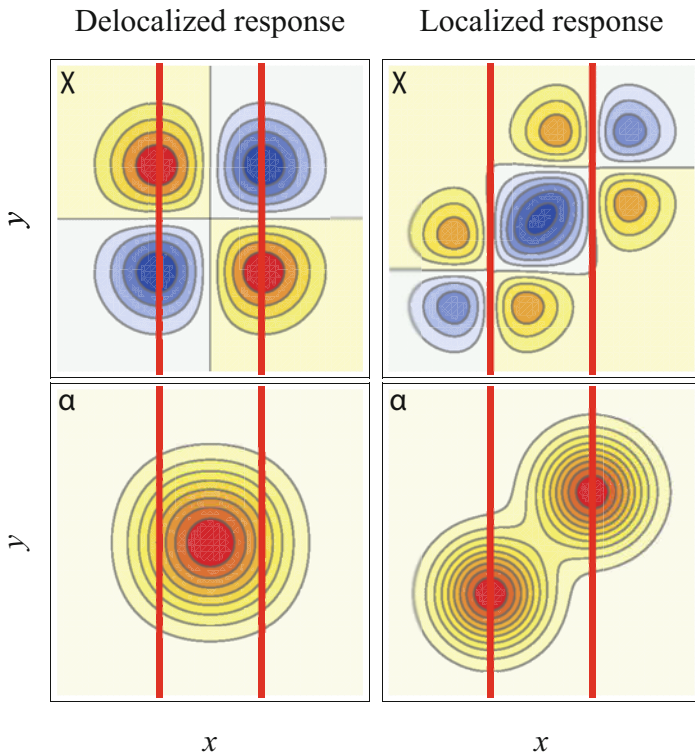
**Fig. 2** Density response function, $\chi(x, y)$, compared to nonlocal dipole polarizability, $\alpha(x, y)$. The figure considers two different model one-dimensional two-particle systems with particle coordinates $x$ and $y$. The left and right systems have a delocalized and localized response, respectively. The columns encode the response of the two systems in two different but equivalent ways – as a density response function (top), $\chi(x, y)$, and as a nonlocal dipole polarizability (bottom), $\alpha(x, y)$. The red and blue colors correspond to positive and negative values. The red lines denote the positions of the two responding Gaussian charge densities on the $x$-axis

nondissipative and dissipative parts of the response, respectively. The frequency dependence of the imaginary part of the polarizability encodes the full optical (electromagnetic) absorption spectrum. This is equivalent to knowing the full energy spectrum of the corresponding Hamiltonian, which is a much harder problem than calculating the ground-state energy. Accordingly, the ACFD formula for the electron correlation energy contains the polarizability only under the integral sign over all frequencies, and it is sufficient to model the spectrum only such that its sum total is represented accurately, which is conveniently achieved with the so-called Wick rotation,

$$\int_0^\infty \mathrm{d}u \, \mathrm{Im} \, \alpha(u) = \int_0^\infty \mathrm{d}u \, \alpha(\mathrm{i}u) . \tag{9}$$

In contrast to the full absorption spectrum, Im $\alpha(u)$, which is a complicated and highly system-dependent function of the frequency, the imaginary-axis part of $\alpha$, $\alpha(iu)$, is a real-valued monotonically decreasing function, which has the same general shape for all kinds of systems. Together with the localized nature of the polarizability, this justifies the use of the harmonic oscillator to model the frequency dependence of the polarizability of more complex electronic systems. The polarizability of a harmonic oscillator is a simple function with two parameters, the static polarizability, $\alpha_0$, and the resonance frequency, $\omega$,

$$\boldsymbol{\alpha}^{\mathrm{HO}}(iu) = \frac{q^2\omega^2}{m\left(\omega^2 + u^2\right)} \equiv \frac{\boldsymbol{\alpha}_0\omega^2}{\omega^2 + u^2} \tag{10}$$

Consider a system of nonoverlapping bodies of electronic matter interacting via the usual electronic Coulomb force (described equivalently either by the Coulomb or dipole potential). Putting aside the internal structure of these bodies for now, their interaction energy, $E_{\mathrm{int}}$, consists of the electrostatic part (including induction) and the (long-range) electron correlation part, $E_{\mathrm{c}}$ – the vdW energy. The ACFD formula for $E_{\mathrm{c}}$ derives its name from two parts: first, the correlations in the electronic fluctuations can be expressed in terms of the response of the bodies to an external field via the so-called *fluctuation–dissipation* theorem (Landau and Lifschitz 1980, sec. §124). Second, $E_{\mathrm{int}}$ is the difference in the total energy of the system when the Coulomb potential between the interacting bodies is switched on ($\lambda = 1$) and switched off ($\lambda = 0$). This difference can be alternatively expressed as an integral over from 0 to 1 while keeping the system at its ground state for all $\lambda$, in this way *adiabatically connecting* the noninteracting and interacting cases. Putting these two parts together, the ACFD theorem gives $E_{\mathrm{c}}$ in terms of the total polarizability of the nonoverlapping subsystems,

$$-\frac{1}{2\pi}\int_0^\infty du \iint d\mathbf{r}d\mathbf{r}' \int_0^1 d\lambda \operatorname{Tr}\left[\boldsymbol{\alpha}\left(\mathbf{r}, \mathbf{r}', iu; \lambda\right) \mathbf{T}^{\mathrm{int}}\left(\mathbf{r}, \mathbf{r}'\right)\right] \tag{11}$$

where $\mathbf{T}^{\mathrm{int}}(\mathbf{r}, \mathbf{r}')$ is set to zero for $\mathbf{r}, \mathbf{r}'$ from the same subsystem.

## 4.3    Range Separation in Density Functional Theory

The distinction between the short-range and long-range parts of the electron correlation energy becomes blurred in realistic systems at equilibrium geometries, where the overlaps between (or within) the interacting subsystems cannot be neglected. Some ab initio electronic-structure methods can treat both parts on an equal basis, but those are often computationally demanding and not applicable to large systems. It is therefore more common that the short-range and long-range parts are treated with different models, which must be somehow seamlessly joined, to avoid both omitting or double counting of some midrange part of the correlation.

Notably, one of the most successful of such approaches is the combination of the density functional theory (DFT) in semilocal approximation for the short-range correlation and explicit vdW models for the long-range correlation.

The ACFD formula can be derived within the density functional for the total correlation energy of a general electronic system (see Callen and Welton 1951; Parr and Yang 1989, Sect. 8.6; Kohn et al. 1998), not only for the interaction energy of nonoverlapping bodies as in the previous section. In this form, it yields the exchange–correlation (XC) energy, $E_{xc}$, the exchange part stemming from the antisymmetry of the electronic wave function,

$$
\begin{aligned}
E_{xc} &= -\frac{1}{2\pi} \int_0^\infty du \iint d\mathbf{r} d\mathbf{r}' \int_0^1 d\lambda \, \chi \left( \mathbf{r}, \mathbf{r}', iu; \lambda \right) v \left( \mathbf{r}, \mathbf{r}' \right) \\
&= -\frac{1}{2\pi} \int_0^\infty du \iint d\mathbf{r} d\mathbf{r}' \int_0^1 d\lambda \, \mathrm{Tr} \left[ \boldsymbol{\alpha} \left( \mathbf{r}, \mathbf{r}', iu; \lambda \right) \mathbf{T} \left( \mathbf{r}, \mathbf{r}' \right) \right]
\end{aligned} \tag{12}
$$

Within time-dependent DFT, the true response functions can be further expressed in terms of the response functions of the KS noninteracting system, corresponding to $\lambda = 0$, as a Dyson screening equation,

$$
\chi \left( iu; \lambda \right) = \chi \left( iu; 0 \right) + \chi \left( iu; 0 \right) \left( v + f_{xc} \left( iu; \lambda \right) \chi \left( iu; \lambda \right) \right) \tag{13}
$$

where $f_{xc}$ is the so-called XC kernel, in general an unknown system-dependent nonlocal function. In approximate vdW models, the XC effects contained in the kernel are usually incorporated implicitly into the effective polarizability,

$$
\boldsymbol{\alpha} \left( iu; \lambda \right) \approx \boldsymbol{\alpha}^{\mathrm{eff}} \left( iu \right) + \boldsymbol{\alpha}^{\mathrm{eff}} \left( iu \right) \mathbf{T} \boldsymbol{\alpha} \left( iu; \lambda \right) \tag{14}
$$

The XC energy can be formally divided into a short-range (sr) and long-range (lr) part by separating the double spatial integral in Eq. 12 into two parts using some range-separating function, $f$, which should decay at least exponentially fast for large distances,

$$
\begin{aligned}
\iint d\mathbf{r}_1 d\mathbf{r}_2 &= \iint d\mathbf{r}_1 d\mathbf{r}_2 \left( 1 - f \left( \mathbf{r}_1, \mathbf{r}_2 \right) \right) + \iint d\mathbf{r}_1 d\mathbf{r}_2 f \left( \mathbf{r}_1, \mathbf{r}_2 \right) \\
&\equiv \iint_{sr} d\mathbf{r}_1 d\mathbf{r}_2 + \iint_{lr} d\mathbf{r}_1 d\mathbf{r}_2
\end{aligned} \tag{15}
$$

In practice, the short-range part is calculated via some semilocal XC functional, for which the corresponding $f$ is system-dependent and in general unknown. On the other hand, most vdW models have $f$ explicitly built in, and the aim is to find explicit $f$ for the vdW method that matches the implicit $f$ of a given XC functional. This is in general an unsolved problem, and most current approaches resort to a varying degree of empiricism.

The most radical difference in the range separation to semilocal DFT and vdW methods runs along the border between systems with uniform and nonuniform electron density. The vdW force between atomic bodies held together by covalent, ionic, or metallic binding is always caused by the long-range electron correlation, but not all effects of the long-range correlation are considered to be a vdW force.

In metals, the electrons from the nonconducting bands are localized on atoms, which form nonuniform islands in the sea of approximately uniform electron density of the conducting electrons (Tao et al. 2010). Here, the long-range correlation between the conducting electrons contributes to the metallic binding. In nonmetals, however, all electrons are nonconducting, the electron density is nowhere uniform, and long-range correlation is mostly associated with vdW interactions. The electronic structure within a single uniform subsystem differs qualitatively in many aspects from that in a nonuniform system. In a uniform system, the exchange effects, the KS density response function, and the XC kernel decay only algebraically with distance (they are long-ranged) as a result of the conducting electrons, whereas they decay exponentially (they are short-ranged) in nonuniform systems (Ge and Lu 2015). (The true density response function decays algebraically in both cases because of electron correlation.) Correspondingly, semilocal and hybrid XC functionals capture both short-range and long-range part of the XC energy in uniform systems but only the short-range part in the nonuniform systems. The vdW interactions can be therefore associated with all long-range electron correlation except for that between conducting electrons within a single uniform subsystem, which is fortunately covered by semilocal and hybrid density functionals. The nonuniform situations include interactions between conducting electrons in disjoint metallic bodies; interactions of conducting electrons with localized electrons, either in the same metallic body or in other bodies; and all interactions between localized electrons.

The consequences of the differences between uniform and nonuniform systems for the range separation can be summarized as follows:

$$
\begin{array}{l}
\text{uniform :} \\
E_{\mathrm{xc}} \quad = \quad \overbrace{\underbrace{E_{\mathrm{x}}^{\mathrm{sr}} + E_{\mathrm{c}}^{\mathrm{sr}}}_{\text{semilocal/hybrid}} + \underbrace{E_{\mathrm{x}}^{\mathrm{lr}}}_{\approx 0} + \underbrace{E_{\mathrm{c}}^{\mathrm{lr}}}_{\text{vdW}}}^{\text{semilocal/hybrid}} \\
\text{nonuniform :}
\end{array} \tag{16}
$$

With the caveat about the uniform systems, the vdW interactions can then be associated with the long-range XC energy. In this setup, care must be taken about the potential double counting of the long-range XC energy in uniform systems from the semilocal or hybrid functionals and from the long-range ACFD formula. This double counting does not matter in situations when the result of a calculation is an energy difference, such as when calculating the adsorption energy of a molecule on a metal surface. But it may pose a problem in other cases, for instance, when investigating a lattice expansion of a metal.

## 5    Classification of van der Waals Functionals

Most existing models of long-range correlation can be described in terms of various approximations to the range-separated effective-polarizability version of the ACFD formula (Eqs. 12 and 14) (Hermann et al. 2017). One of them is the already

discussed local representation of the effective polarizability. Two other general and common approximations are spatial coarse-graining of the system and truncation of the many-body expansion of the vdW energy.

## 5.1 Coarse-Graining of Response Functions

Given a set of functions, $w_p(\mathbf{r})$, that partition space into fragments, $\sum_p w_p(\mathbf{r}) \equiv 1$, and respective centers of the fragments, $\mathbf{R}_p$, each spatial function or operator, such as the dipole polarizability, can be represented as a sum over the partitioned components, $\boldsymbol{\alpha}_{pq}$, which can be in turn expanded in the basis of solid harmonics (multipole expansion), $\alpha_{pq,ll'mm'}$, around the centers (Stone 2013),

$$\boldsymbol{\alpha}\left(\mathbf{r}, \mathbf{r}', u\right) = \sum_{pq} w_p\left(\mathbf{r}\right) w_q\left(\mathbf{r}'\right) \boldsymbol{\alpha}\left(\mathbf{r}, \mathbf{r}'u\right) \equiv \sum_{pq} \boldsymbol{\alpha}_{pq}\left(\mathbf{r}, \mathbf{r}', u\right) \rightarrow \alpha_{pq,ll'mm'} \tag{17}$$

(Here, $l$, $l'$ start from 1, because the expanded quantity is a tensor. The corresponding expansion of the scalar density response function, $\chi$, would start from $l = l' = 0$.) The dipole potential is expanded correspondingly. Unlike the Fourier transformation, the multipole expansion is not invertible, but like the Fourier transformation, it introduces a correspondence between spatial integrals and infinite sums,

$$\mathbf{P}\left(\mathbf{r}, u\right) = -\int d\mathbf{r}' \boldsymbol{\alpha}\left(\mathbf{r}, \mathbf{r}', u\right) \mathbf{E}\left(\mathbf{r}', u\right) \iff \mathbf{P}_{p,lm}(u)$$
$$= -\sum_{q,l'm'} \alpha_{pq,ll'mm'}(u) E_{q,l'm'} \tag{18}$$

The motivation for this multipole reformulation is that because both $\mathbf{T}_{\text{eff}}$ and $\mathbf{T}_{\text{lr}}$ are long-ranged and their moments decay increasingly faster for higher $l$'s; all the matrix multiplications (infinite sums) converge quickly and can be approximated well by finite sums.

The feasibility of the coarse-graining and multipole expansions is dictated by the choice of the fragments and the response properties of the system. In a nonuniform system, the nonlocal effective polarizability is exponentially localized on atoms, and atom-centered fragments are a natural basis of a quickly converging multipole expansion. In a uniform system, the effective polarizability is long-ranged, diffused, and there are no natural centers for the multipole expansion, leading to large higher moments and slow convergence or even divergence of the expansion. In principle, this problem is mitigated in combination with the KS-DFT, because the long-range XC energy within the uniform systems is captured by the semilocal or hybrid functional, and the multipole convergence of the correlation energy due to the interaction with a separate uniform or nonuniform system is helped by larger

separations between the fragments. But such an interplay is not well understood, and none of the coarse-grained models discussed in this chapter take advantage of this cancellation.

## 5.2 Truncation of Many-Body Expansion

After inserting the Dyson screening equation into the ACFD formula, the coupling-constant integration can be carried out analytically either when using the effective-polarizability formulation or by approximating the XC kernel with some form that depends on explicitly. In both cases, the $\lambda$-integration results in a logarithm expression such as

$$E_{xc} \approx -\frac{1}{2\pi} \int_0^\infty du \iint d\mathbf{r} d\mathbf{r}' \, \mathrm{Tr} \left[ \ln \left( 1 - \boldsymbol{\alpha} \left( \mathbf{r}, \mathbf{r}', iu; \lambda = 0 \right) \mathbf{T} \left( \mathbf{r}, \mathbf{r}' \right) \right) \right] \qquad (19)$$

The operator logarithm is defined as an infinite series, and writing it out explicitly in terms of individual orders leads to a many-body decomposition of the XC energy,

$$\begin{aligned}
E_{xc} = &\frac{1}{2\pi} \int_0^\infty du \iint d\mathbf{r} d\mathbf{r}' \, \mathrm{Tr} \left[ \boldsymbol{\alpha} \left( \mathbf{r}, \mathbf{r}', u; 0 \right) \mathbf{T} \left( \mathbf{r}, \mathbf{r}' \right) \right] \\
&- \frac{1}{4\pi} \int_0^\infty du \iiint \int d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' d\mathbf{r}''' \, \mathrm{Tr} \left[ \boldsymbol{\alpha} \left( \mathbf{r}, \mathbf{r}', u; 0 \right) \mathbf{T} \left( \mathbf{r}', \mathbf{r}'' \right) \right. \\
&\left. \boldsymbol{\alpha} \left( \mathbf{r}'', \mathbf{r}''', u; 0 \right) \mathbf{T} \left( \mathbf{r}''', \mathbf{r} \right) \right] + \frac{1}{6\pi} \cdots + \ldots
\end{aligned} \qquad (20)$$

The term "many-body" is best motivated in the coarse-grained models where the individual terms correspond to interactions between increasing number of fragments (bodies). (The order does not necessarily correspond to the number of bodies. At fourth order, for instance, some terms are a back-and-forth interaction between two bodies.) When constructed from the bare KS polarizability, the first order evaluates to the exact exchange, and the correlation energy starts at the second order. With any local approximation for the bare polarizability, the first term evaluates identically to zero, and the formula gives only the correlation energy. The long-range part of the second term is the leading term for vdW interactions and the basis of all nonlocal correlation functionals and coarse-grained pairwise methods discussed below. The third term corresponds to the Axilrod–Teller–Muto (ATM) three-body potential (Axilrod and Teller 1943; Mutō 1943) when coarse-grained to atoms.

## 5.3 Random-Phase Approximation

The approximations to the ACFD formula that are fully many-body and not coarse-grained can be based on the bare KS density response function. Because the KS

density response function can be calculated directly from the KS orbitals via the so-called Adler–Wiser formula, these approximations are usually formulated and evaluated in the $\chi \upsilon$-representation rather than the $\boldsymbol{\alpha T}$-representation. Furthermore, because the KS response has a well-defined short-range structure, this construction allows the evaluation of the total XC energy, not only its long-range part, so the use of these methods goes beyond long-range correlation energy. Here, we discuss the methods from the perspective of vdW interactions.

The simplest of these methods is the RPA itself, which amounts to setting the XC kernel in Eq. 13 to zero (Ren et al. 2012). The omitted XC kernel is short-ranged in nonuniform systems, but its omission influences both short-range and long-range correlation energy, because the short-range XC effects still influence the total polarizability of the system, which is manifested in the long-range correlation energy via the ACFD formula. As a result, although RPA does not suffer from any systematic errors in the long-range correlation energies, the overall accuracy is often worse than that of the many effective models discussed below (Olsen and Thygesen 2013b). This is further amplified in vdW systems in equilibrium geometries, where the short-range XC energy also contributes to the total interaction energy. Attempts at improvement go both ways, replacing the short-range correlation energy with a better model than RPA and improving the effective polarizability.

Kurth and Perdew (1999) suggested to correct the short-range correlation energy from RPA with that from a semilocal XC functional, in what they called the RPA+ method. Rather than explicitly range-separating the ACFD expression, RPA+ removes the RPA short-range part by subtracting correlation energy from a specially designed semilocal correlation functional, $E_c^{\text{GGA@RPA}}$, and reintroduces it back with a standard semilocal functional, $E_c^{\text{GGA}}$.

$$E_c^{\text{RPA+}} = E_c^{\text{RPA}} - E_c^{\text{GGA@RPA}} + E_c^{\text{GGA}} \tag{21}$$

$E_c^{\text{GGA@RPA}}$ is constructed in a similar way as standard functionals, but its uniform part is parameterized to reproduce the RPA energy of the electron gas rather than the true energy. RPA+ attempts to fix the short-range correlation energy of RPA, but the long-range part is unchanged, so the vdW force remains the same, and it is only the interaction due to electron-density overlap, which occurs at equilibrium, that can be possibly improved. Furthermore, the range separation in RPA+ is unsystematic in the sense that there is no guarantee that $E_c^{\text{GGA@RPA}}$ and $E_c^{\text{GGA}}$ have the same effective range.

Toulouse et al. (2004) formulated a range-separated version of the KS scheme, in which the XC functional is designed from the beginning to treat only the short-range part of the electron correlation. This leads to an alternative range separation of the ACFD formula, in which $\boldsymbol{\alpha}(\lambda)$ is not the polarizability of the wave function that minimizes $\langle \Psi | \widehat{T} + \lambda \widehat{V} | \Psi \rangle$ but rather of one that minimizes $\langle \Psi | \widehat{T} + \lambda \widehat{V}_{\text{lr}} | \Psi \rangle$ (Toulouse et al. 2009). In this scheme, the RPA of the Dyson-like equation results in a model in which the effective polarizability is still equal to the bare KS polarizability, like in normal RPA, but the effective dipole operator is only the

long-range part of the full operator. The underlying assumption then is that the dipole operator and the XC kernel partially cancel out at short range, giving a different estimate of the effective polarizability than normal RPA. This is supported by numerical evidence on select small systems. A similar scheme, proposed earlier by Kohn et al. (1998), also uses a range-separated version of the KS scheme, but instead of obtaining the true polarizability at the RPA level, $\chi(\lambda)$ is obtained for each $\lambda$ by explicitly perturbing the corresponding $\lambda$-scaled system with electric field.

A straightforward way to improve the RPA is to devise approximate XC kernels, which improves the short-range behavior of the polarizability and hence both short-range and long-range correlation energies. Extending the LDA to the time domain, the adiabatic LDA (ALDA) assumes that the XC kernel has no memory, leading to a frequency-independent local XC kernel. Unlike LDA, which is exact for the uniform electron gas (UEG), ALDA does not give the true XC kernel of the UEG (which is nonlocal in both time and space) and violates several known properties of the true XC kernel. Despite that, it is a useful approximation in TD-DFT calculations when one is interested only in a certain range of the frequency spectra. Still, it turns out not to be a good approximation in the ACFD formula, where it gives worse results than the absent XC kernel of the RPA (Lein et al. 2000).

Olsen and Thygesen (2012) constructed a correction to ALDA by fixing its large-$\mathbf{q}$ (short-range) behavior in the UEG to better reproduce the known exact behavior. Taking this renormalized ALDA (rALDA) kernel, transforming back to real space, and using the mean density in two points as the corresponding uniform density, this procedure gives a universal XC kernel. This construction is computationally no more demanding than RPA but improves upon RPA in almost every case tested (Olsen and Thygesen 2013a, 2014). The rALDA XC kernel gives a more realistic short-range screening of the bare KS polarizability, resulting in more accurate long-range correlation energies and better description of vdW systems.

A different path toward improving the accuracy of RPA can be taken using the many-body perturbation (MBPT) theory. This is possible because, as Gell-Mann and Brueckner (1957) showed, yet another equivalent definition of RPA is via a certain subset of Feynman diagrams, the so-called ring diagrams. Summing different subsets of the diagrams similar to those corresponding to RPA then leads to different RPA-like models and sometimes confusing terminology, when a certain modification of the XC kernel in RPA is equivalent to adding additional terms to the RPA XC energy that do not seem to be related to RPA (Scuseria et al. 2008; Jansen et al. 2010; Ángyán et al. 2011).

The second-order Møller–Plesset correlation energy (MP2) consists of the Coulomb direct and exchange terms, of which only the former is long-ranged. In this context, RPA can be understood as the sum of all MP2-like direct terms (ring diagrams) in the infinite MBPT expansion. Similarly, the MP2 exchange can be renormalized by replacing one of the Coulomb interactions with the RPA sequence of ring diagrams, leading to the second-order screened exchange (SOSEX). Furthermore, unlike in the Møller–Plesset perturbation theory, where the first order is guaranteed to be zero, single-electron excitations contribute to the XC energy in the MBPT based on KS orbitals. Combining RPA, SOSEX,

and RPA-renormalized single-excitation correction then results in the renormalized second-order perturbation theory (rPT2) (Ren et al. 2011, 2013). Although the MP2 exchange term is short-ranged, the renormalization in SOSEX is long-ranged, and the long-range correlation energy of rPT2 is different from that of RPA. The combined improvements of the short-range and long-range XC energy in rPT2 compared to RPA lead to improved accuracy in vdW binding energies.

## 5.4    Nonlocal van der Waals Density Functionals

The models of long-range correlation energy discussed in this section are in the class of approximations to the ACFD formula that truncate the many-body expansion at second order, but do not do any spatial coarse-graining. This leads to XC functionals that are characterized by nonlocal dependence of the XC energy density on the electron density via some nonlocal kernel, $\Phi$,

$$
\begin{aligned}
E_{\text{c}}^{\text{nl-df}} &= -\frac{1}{2} \int d\mathbf{r} d\mathbf{r}' n\left(\mathbf{r}\right) n\left(\mathbf{r}'\right) \Phi\left[n\right]\left(\mathbf{r}, \mathbf{r}'\right) \\
&= -\int d\mathbf{r}\, n\left(\mathbf{r}\right) \int d\mathbf{r}' \frac{1}{2} n\left(\mathbf{r}'\right) \Phi\left[n\right]\left(\mathbf{r}, \mathbf{r}'\right)
\end{aligned}
\tag{22}
$$

The effective polarizability is approximated with a local isotropic polarizability,

$$
\boldsymbol{\alpha}_{\text{eff}}\left(\mathbf{r}, \mathbf{r}', u\right) \approx \mathbf{I}\alpha_{\text{eff}}\left(\mathbf{r}, u\right) \delta\left(\mathbf{r} - \mathbf{r}'\right)
\tag{23}
$$

This results in the first-order term in the many-body expansion being zero, which means that such a functional cannot capture any exchange energy, which is intentional, since the nonlocal functionals are designed to capture only the long-range correlation energy. The locality of the effective polarizabilities reduces two of the four integrals in the second-order term, and the isotropy allows to take the polarizabilities out of the trace in Eq. 20.

A general form of the local effective polarizability used in many models is obtained from the polarizability of a harmonic oscillator by setting the ratio of the charge and mass to that of an electron, $q/m = 1$, and substituting the electron density for the charge,

$$
\alpha_{\text{tot}}^{\text{HO}}\left(\mathrm{i}u\right) = \frac{q^2/m}{\omega^2 + u^2} \rightarrow \alpha_{\text{eff}}\left[n\right]\left(\mathbf{r}, \mathrm{i}u\right) = \frac{n\left(\mathbf{r}\right)}{\omega_{\text{eff}}^2\left[n\right]\left(\mathbf{r}\right) + u^2}
\tag{24}
$$

Besides the obvious reason of modelling electrons, the charge–mass ratio of one is motivated by the $f$-sum rule for an electronic system that dictates that $\alpha_{\text{tot}}(\mathrm{i}u) \rightarrow N/u^2$ ($N$ is the number of electrons), which the form above automatically satisfies. (Strictly speaking, this is not necessary, because the rule does not need to be satisfied in any local form, and furthermore, the local effective polarizability is not supposed to integrate to the total polarizability without any long-range coupling.

However, the local form is a straightforward way to satisfy the global rule.) The local effective resonance frequency, $\omega_{\text{eff}}^2$, can be in general any functional of the electron density but is often approximated semilocally.

The approximated ACFD formula can then be recast in the form of a nonlocal density functional, where the nonlocal kernel is a functional of the effective resonance frequency and some (so far unspecified) range-separating function,

$$
E_{\text{c,lr}} \approx -\frac{1}{2} \iint \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' n\left(\mathbf{r}\right) n\left(\mathbf{r}'\right) \frac{3}{\pi} \int_0^\infty \mathrm{d}u
$$
$$
\frac{1}{\omega_{\text{eff}}^2\left[n\right]\left(\mathbf{r}\right) + u^2} \frac{1}{\omega_{\text{eff}}^2\left[n\right]\left(\mathbf{r}'\right) + u^2} \frac{f\left(\mathbf{r}, \mathbf{r}'\right)}{\left|\mathbf{r} - \mathbf{r}'\right|^6}
\tag{25}
$$

The asymptotic behavior of the long-range correlation energy calculated in this way is fully specified by $\omega_{\text{eff}}$.

The first general functional of this form, referred to simply as the vdW density functional (vdW-DF), was developed by Dion et al. (2004). Although the derivation of vdW-DF starts from the ACFD formula, it follows quite a different direction than the framework in this chapter, and most of the approximations along the way are done in reciprocal space, until everything is transformed back to real space in the end. However, the final result can still be cast in the form of Eq. 25. The effective resonance frequency in the vdW-DF is constructed from a GGA-type XC energy density. The first equality is motivated by using $\omega_{\text{eff}}^2$ to calculate the XC energy of a slowly varying electron gas via the ACFD formula. The particular choice of the semilocal approximation to the XC energy density is rather arbitrary and completely independent of the semilocal functional potentially used to complete the vdW-DF at short range.

A serious disadvantage of the vdW-DF in light of other long-range correlation models is that its range-separating function is fixed by the underlying theory. Because of the construction in the reciprocal space, the parameter $A$ appears both in the effective resonance frequency and the range-separating function. Since the asymptotic behavior of any nonlocal functional depends only on $\omega_{\text{eff}}$, not the range-separating function, the parameter $A$ is essentially fixed, and there is no remaining freedom in the range-separating function that could be adjusted for a particular choice of a short-range semilocal functional in a full KS-DFT calculation.

The form of the range-separating function is complex due to the reciprocal-space formulation, but there are two underlying physical motivations for it. When the two oscillators given by the resonance frequencies $\omega_{\text{eff}}$ are close to each other such that their ground-state wave functions would overlap, the underlying model does not work anymore, the corresponding part of the XC energy must be covered by the semilocal functional, and the dipole coupling must be damped. This is effectively achieved by increasing the resonance frequency as $k^2$ in the reciprocal space. The second damping mechanism is that the nonlocal functional must evaluate to zero for the uniform electron gas, whose long-range correlation energy is already covered by a semilocal or a hybrid functional. This forces the range-separating function to

negative values at short range, to counterbalance the attractive contribution from the long range.

The complex form of vdW-DF was simplified in the VV09 functional, which used a substantially different form of $\omega_{\mathrm{eff}}$,

$$\omega_{\mathrm{VV}}^2 [n] (\mathbf{r}) = \frac{4\pi}{3} n (\mathbf{r}) + C \frac{|\nabla n|^4}{n^4} \tag{26}$$

Here, $4\pi n$ is the resonance frequency of the macroscopic (small-$\mathbf{q}$ limit) plasmon fluctuations of the uniform electron gas. The factor of 1/3 comes from the Clausius–Mossotti relation between the microscopic local polarizability and the macroscopic dielectric function. The density-gradient term is a local model of a bandgap obtained from considering the behavior of the electron density in the density tail of a finite system. The range-separating mechanism of VV09 is still constructed in reciprocal space. The latest attempt at a simplified formulation of the vdW-DF, named VV10, was constructed entirely in real space (Vydrov and Van Voorhis 2010). Both the resonance frequency and range-separating function of Eq. 25 have a simple form in VV10. The former is the same as in VV09, and the latter is constructed using the same mechanism of reduced polarizabilities of overlapped oscillators as in the original vdW-DF but in real space.

## 5.5    Pairwise Interatomic Models

The oldest approaches to fix the missing long-range electron correlation in HF or semilocal KS-DFT calculations are of the interatomic pairwise form,

$$E_{\mathrm{c,lr}} \approx -\frac{1}{2} \sum_{pq} C_{6,pq} \frac{f (\mathbf{R}_p, \mathbf{R}_q)}{|\mathbf{R}_p - \mathbf{R}_q|^6} \tag{27}$$

Here, $f$ is some range-separating (damping) function, $\mathbf{R}_q$ are the atom coordinates, and the so-called dispersion coefficients, $C_{6,pq}$, determine the asymptotic interaction between two atoms. This type of interatomic potential has origin in empirical force fields dating back to the Lennard–Jones potential, even before it was clear that the correct leading term of the vdW force is $1/R^6$. In the context of electronic-structure methods, it was first used by Hepburn et al. (1975) to correct interaction curves of rare-gas dimers from HF calculations. This approach was later extended to molecules and KS-DFT calculations, and the $C_6$ coefficients were extended to a wider range of systems (Halgren 1992; Mooij et al. 1999; Elstner et al. 2001; Wu and Yang 2002). Grimme (2004) then presented a parametrization of $C_6$ and $f$, termed DFT-D ("D" for dispersion), that could in principle be applied to any molecule or solid, in combination with any XC functional. This marked a start of routine addition of the long-range correlation energy to semilocal KS-DFT calculations.

The pairwise interatomic model of Eq. 27 can be obtained as a coarse-grained truncated approximation to the ACFD formula. The derivation follows the same course of second-order truncation and local approximation to the effective polarizability as nonlocal vdW density functionals but starting from the coarse-grained multipole-expanded ACFD formula (see Eq. 18),

$$E_{c,lr} \approx -\frac{1}{4\pi} \int_0^\infty du \, \mathrm{Tr}_{p,lm} \left( \boldsymbol{\alpha}_{\mathrm{eff}} (iu) \, \mathbf{T}_{lr} \boldsymbol{\alpha}_{\mathrm{eff}} (iu) \, \mathbf{T}_{lr} \right) \tag{28}$$

Here, the trace is over multipole moments and fragments, which are chosen to be atoms in most cases. (In this context, the formal definition of an atom in a molecule is given by some partition function.) Approximating the local effective polarizability as isotropic, $\alpha_{\mathrm{eff}, pll'mm'} = \delta_{ll'}\delta_{mm'}\alpha_{\mathrm{eff}, pl}$, the formula is reduced as in the case of nonlocal vdW XC functionals. The standard pairwise formula of Eq. 27 is recovered by keeping only the lowest dipole–dipole term ($l = l' = 1$, $K_{11} = 6$), where the expression for the corresponding dispersion coefficient is called the Casimir–Polder integral,

$$C_{6, pq} = \frac{3}{\pi} \int_0^\infty du \, \alpha_{\mathrm{eff}, p1} (iu) \, \alpha_{\mathrm{eff}, q1} (iu) \tag{29}$$

Some pairwise methods are formulated directly in terms of the dispersion coefficients, not the underlying polarizabilities, in which case approximate combination rules for calculating unknown heteronuclear coefficients from known homonuclear coefficients are useful. Such rules can be derived from the Casimir–Polder integral using some model polarizability. An often used rule is obtained from the harmonic-oscillator model,

$$C_{6, pq}^{\mathrm{HO}} = \frac{3}{\pi} \int_0^\infty du \frac{\alpha_p(0)\omega_p^2}{\omega_p^2 + u^2} \frac{\alpha_q(0)\omega_q^2}{\omega_q^2 + u^2} = \frac{2C_{6, pp}C_{6, qq}}{C_{6, pp}\frac{\alpha_q(0)}{\alpha_p(0)} + C_{6, qq}\frac{\alpha_p(0)}{\alpha_q(0)}} \tag{30}$$

Using the single-pole polarizability of the harmonic oscillator in situations where the true spectrum is more complex, such as in the equation above, is called the Unsöld (1927) approximation.

The models of Grimme, mentioned above, are different from the rest discussed in this section in that they are formulated only in terms of the geometry of a molecule, $\{\mathbf{R}p\}$, not the electron density. This makes them straightforwardly useful even for empirical short-range electronic models that do not produce any electronic density, but at the same time, it makes it much harder to achieve truly general models, because the electron density encodes much useful information about the system.

The first version of DFT-D used fixed homonuclear $C_6$ coefficients, the combination of Eq. 30 with all polarizability ratios set to 1 and a range-separating function constructed from vdW radii that did not go to 1 in infinity (Grimme 2004). The second version was a numerical reparametrization of the first one with a changed combination rule, which set the polarizability ratios equal to those of

the $C_6$ coefficients (Grimme 2006). In the first and second version, the atomic $C_6$ coefficients do not depend on the molecular environment, which is a crude approximation. The third version was an improvement in several regards (Grimme et al. 2010). The range separation was modified to obey the correct asymptotic behavior. An elementary dependence of the $C_6$ coefficients on the environment was included via geometrical factors estimating the coordination number of an atom. The dipole–quadrupole term ($l = 1$, $l' = 2$) was included, and a three-atom correction was suggested, which is the third-order triple-dipole term in the logarithm expansion of the coarse-grained ACFD formula. The corresponding dispersion coefficients, $C_8$ and $C_9$, are obtained by combination rules similar to those for the $C_6$ coefficient.

Soon after the first version of DFT-D and in stark contrast to it, Becke and Johnson (2005b) developed a method to calculate $C_6$ coefficients from first principles, using an approximation to the polarizability based on the dipole moment of the XC hole of the HF model, the exchange-hole dipole method (XDM). Their initial derivation was rather heuristic, with a wrong prefactor, but the final result can be in fact obtained directly from the Casimir–Polder integral using the fluctuation–dissipation theorem for the density response function and the Unsöld approximation (Ángyán 2007; Heßelmann 2009; Ayers 2009). A semilocal approximation to the XC hole by Becke and Roussel (1989) works as well as that from the HF model and with the additional benefit of reduced computational complexity (Becke and Johnson 2005a). To formulate a general interatomic pairwise method, the (local) dipole moment of the XC hole is coarse-grained using the partitioning scheme devised by Hirshfeld (1977). In this scheme, the atomic partition functions, $w_p$, are constructed from radially averaged electron densities of isolated atoms, $n^{\text{free}}$,

$$w_p^{\text{Hirsh}}(\mathbf{r}) = \frac{n_p^{\text{free}}(|\mathbf{r} - \mathbf{R}_p|)}{\sum_q n_q^{\text{free}}(|\mathbf{r} - \mathbf{R}_q|)} \tag{31}$$

The corresponding static dipole polarizabilities of the atomic fragments are calculated from free-atom dipole polarizabilities, assuming that they scale linearly with the Hirshfeld measure of a volume (Hirshfeld volume),

$$\alpha_{p1}(0) = \alpha_{p1}^{\text{free}}(0) \frac{V_p^{\text{Hirsh}}[n]}{V_p^{\text{Hirsh}}[n^{\text{free}}]} \tag{32}$$

$$V_p^{\text{Hirsh}}[n] = \int d\mathbf{r}\, n(\mathbf{r})\, w_p^{\text{Hirsh}}(\mathbf{r})\, |\mathbf{r} - \mathbf{R}_p|^3 \tag{33}$$

The fragment $C_6$ coefficients are then calculated from the fragment polarizabilities and coarse-grained XC hole dipole moment. The harmonic-oscillator combination rule is used to get the rest of the $C_6$ coefficients. The XDM can be extended to higher multipole dispersion coefficients by calculating higher multipole moments of the XC hole polarization around each atomic center (Becke and Johnson 2006; Johnson and Becke 2006).

A simple yet accurate interatomic pairwise method was developed by Tkatchenko and Scheffler (2009) (TS), who extended the free-atom scaling approach to all the atomic parameters, including the $C_6$ coefficients and the vdW radii, and thus formulated the calculation of interatomic pairwise vdW interactions into a true density functional. Assuming that the excitation energies of the atoms are independent of the volume, the Unsöld approximation and the Casimir–Polder integral dictate that the $C_6$ coefficients scale with the second power of the Hirshfeld volume ratio,

$$C_{6,pq} = C_{6,pq}^{\text{free}} \left( \frac{V_p^{\text{Hirsh}}[n]}{V_p^{\text{Hirsh}}\left[n^{\text{free}}\right]} \right)^2 \tag{34}$$

The free-atom reference values may not be the most effective choice in metals and some solids, whose electron density is often substantially different from the superposition of free-atom densities. Zhang et al. (2011) and Ruiz et al. (2012) used an adapted TS method, where the reference values are obtained from bulk macroscopic dielectric constant.

Sato and Nakai (2009, 2010) developed an atomic pairwise method based on the local effective polarizability functional from the vdW-DF-09 nonlocal functional. A system is described by the local effective polarizability given by the harmonic-oscillator formula with the resonance frequency from Eq. 26. The atomic fragments are defined using the partitioning functions from the scheme by Becke (1988), which is most often used to define atomic radial grids in KS-DFT calculations, but here it is used as an alternative to the Hirshfeld partitioning. The partitioned polarizability is used to calculate a coarse-grained representation of the system via multipole expansion and Casimir–Polder integrals up to the $C_{10}$ coefficient.

Silvestrelli (2008) formulated a pairwise method in which the coarse-grained fragments are not atoms but Wannier functions (WFs) (Marzari et al. 2012). Wannier functions are any set of localized one-electron wave functions that in principle form a complete basis. In finite molecular systems, they are called Boys orbitals. The Wannier functions of conducting and nonconducting electrons are localized algebraically and exponentially, respectively. In the vdW-WF method, each WF is approximated with a single spherically symmetric exponential function that has the same width (second central moment) as the true WF. The polarizability of the approximate WF is calculated with the polarizability functional of Andersson et al. (1996) (ALL). Here, $n_p$ is the electron density of the WF and $k$ is a nonempirical constant. The $C_6$ coefficients between the WFs are calculated from the Casimir–Polder integral, and the range-separating function is the same as in the TS method, with vdW radii of the WFs defined via an electron density cutoff (Silvestrelli et al. 2009). The vdW-WF scheme has two theoretical shortcomings: first, the partitioning of the total electron density is only approximate because of the use of the approximate WFs, and second, the ALL polarizability functional was designed for the total electron density, not one-electron densities.

Coarse-grained methods in which the fragment polarizabilities and $C_6$ coefficients are calculated directly, rather than obtained by explicit partitioning of some

continuous quantity, may be sensitive to a particular choice of the partitioning scheme. This motivated a series of modified Hirshfeld partitioning schemes that should capture better the redistribution of the electron density in a molecule with respect to free atoms. Steinmann and Corminboeuf (2010, 2011) adapted the XDM to use the self-consistent Hirshfeld scheme, which gives a more consistent description of ionic systems (Bultinck et al. 2007). Bučko et al. (2013, 2014) did the same with the TS method. The self-consistent Hirshfeld partitioning uses the same stockholder formula in Eq. 31 as the original scheme, but the reference densities are generalized and depend recursively on the partitioning, leading to equations that need to be solved iteratively (Verstraelen et al. 2012). A common form of the generalized reference densities, used in the modified XDM and TS methods, is a linear combination of free-atom and free-ion densities that maintains the charge of the Hirshfeld-partitioned atomic density. This scheme is complicated by the instability of many isolated anions, which requires addition of auxiliary negative charges, making the partitioning somewhat arbitrary.

## 5.6 Many-Body Dispersion

The fourth class of approximations to the ACFD formula covers nontruncated coarse-grained models. A common theme of all such models is to interpret the Unsöld approximation with its single resonance frequency literally and model a real molecular system as a collection of coupled charged oscillators. The corresponding Hamiltonian describes a system of distinguishable particles characterized by a charge, $q_i$, and a mass, $m_i$, each having its own harmonic potential defined by the resonance frequency, $i$, and a center, $\mathbf{R}_i$, interacting via the Coulomb force,

$$
\begin{aligned}
\widehat{H}_{\text{osc}} = & \sum_i \frac{\widehat{\mathbf{p}}_i^2}{2} + \sum_i \frac{1}{2} m_i \omega_i^2 |\widehat{\mathbf{r}}_i - \mathbf{R}_i|^2 \\
& + \sum_{i<j} q_i q_j \left( \frac{1}{|\widehat{\mathbf{r}}_i - \widehat{\mathbf{r}}_j|} - \frac{1}{|\widehat{\mathbf{r}}_i - \mathbf{R}_j|} - \frac{1}{|\mathbf{R}_i - \widehat{\mathbf{r}}_j|} + \frac{1}{|\mathbf{R}_i - \mathbf{R}_j|} \right)
\end{aligned}
\tag{35}
$$

The centers of the harmonic potentials additionally host a compensating charge of the opposite sign. If the centers are the same as those of the atoms, this Hamiltonian can be interpreted as a very crude approximation to the electronic Hamiltonian, in which all electrons of individual atoms are described by distinguishable psuedoelectrons that move in an effective potential which is the combined result of the nuclear potential and the mean field of the electrons. In particular, any exchange effects and hence charge transfer and delocalization are not considered. Expanding the Coulomb operator in a multipole series and keeping only the dipole term result in dipole-coupled oscillator Hamiltonian,

$$\widehat{H}_{\text{dosc}} = \sum_i \frac{\widehat{\mathbf{p}}_i^2}{2} + \sum_i \frac{1}{2} m_i \omega_i^2 |\widehat{\mathbf{r}}_i - \mathbf{R}_i|^2$$
$$+ \sum_{i<j} q_i q_j \left(\widehat{\mathbf{r}}_i - \mathbf{R}_i\right) \mathbf{T} \left(\mathbf{R}_j - \mathbf{R}_i\right) \left(\widehat{\mathbf{r}}_j - \mathbf{R}_j\right) \tag{36}$$

A useful property of this Hamiltonian is that it can be solved exactly by coordinate transformation into a system of uncoupled quasi-oscillators, which describe different collective oscillations. The ground-state wave function of such a system is then a simple product of the single-oscillator ground-state wave functions, and the ground-state energy is a sum of the single-oscillator ground-state energies, $E_0 = \sum_n \tilde{\omega}_n/2$. Drawing analogy with the RPA, the individual oscillators model the particle-like quasi-electrons in some coarse-grained way, while the coupled oscillations model the wavelike electron oscillations. This Hamiltonian has been used many times to obtain various qualitative properties of long-range electron correlation (Bade 1957; Bade and Kirkwood 1957; Mahan 1965; Lucas 1967; Renne and Nijboer 1967; Donchev 2006) but only recently to formulate general quantitative methods.

The relevance of the dipole-coupled oscillator model to the true electronic system can be derived directly from the coarse-grained ACFD formula by performing the frequency integration analytically (Tkatchenko et al. 2013). When truncated at the dipole term, the approximate long-range correlation energy is then equal to the difference in the ground-state energy between the dipole-coupled oscillators and noninteracting oscillators. The exact equivalence between the dipole-coupled oscillators and the approximated ACFD formula breaks when going beyond the dipole approximation. The effective Hamiltonian derived from the ACFD formula is always bilinear in the interaction and contains one oscillator per each fluctuating moment (dipole, quadrupole, etc.). In contrast, the coupled-oscillator Hamiltonian has always $3N$ coordinates, independent of the degree of the multipole expansion of the Coulomb operator, and the interaction terms above the dipole order are formed from nonlinear combinations of the coordinates, making the Hamiltonian unsolvable in closed form.

The use of the coupled-dipole approach to formulate general methods for the long-range correlation energy was initiated in the many-body dispersion (MBD) model developed by Tkatchenko et al. (2012). MBD reuses the effective dynamic polarizability as approximated in the TS pairwise method and combines it with a physically motivated effective dipole operator. Motivated by the Gaussian shape of the harmonic-oscillator ground-state wave function, the dipole potential in MBD, $\mathbf{T}_{\text{gg}}$, is derived from the screened Coulomb interaction between two Gaussian unit-charge densities (Mayer 2007), with widths derived from the corresponding dipole polarizabilities. In general, the dipole potential in the Dyson equation should be different from that in the ACFD formula and independent of the XC functional used for the short-range part of the correlation energy. To circumvent this obstacle, Ambrosetti et al. (2014b) separated $\mathbf{T}_{\text{gg}}$ into the long-range part and the short-range

remainder. The long-range correlation energy is then calculated in two steps. First, the effective polarizability is screened by the short-range dipole potential via the Dyson equation. Second, the dipole-coupled Hamiltonian is solved with the long-range dipole potential.

Silvestrelli (2013) developed another method inspired by MBD in which the oscillators do not model the response of the atoms but of Wannier functions. This Wannier-based MBD is to the pairwise vdW-WN method as what the range-separated MBD is to the pairwise TS method. Unlike in vdW-WN, here the polarizabilities of the Wannier functions are not calculated using a local polarizability functional but directly from the Hirshfeld volumes of the Wannier functions.

# 6    Applicability in Material Modelling

Most vdW models have been designed following some set of theoretical principles (as is common in DFT), rather than obtained by a straightforward application of systematic approximations (as is common in quantum chemistry). As a result, a careful attention must be paid to the evaluation of the accuracy of the models, to avoid any systematic bias, both within and between different classes of systems, and to know the level of uncertainty in predicted quantities that one may expect. In this regard, the systematic verification of a given DFT+vdW method is usually achieved through comparison against the results of higher-level (more costly and more accurate) theoretical methods or experimental results with sufficient resolution.

In the case of organic molecules and materials, several benchmark sets of binding energies of complexes and lattice energies of molecular crystals have been established that allow for systematic testing of vdW models. For smaller molecules, the reference data have been obtained by high-level correlated methods of quantum chemistry (Jurečka et al. 2006; Řezáč et al. 2011), for larger molecules by diffusion quantum Monte Carlo (Ambrosetti et al. 2014a) and for molecular crystals by extrapolating experimental lattice enthalpies to zero temperature (Otero-de-la-Roza and Johnson 2012; Reilly and Tkatchenko 2013b). Initially, the development of vdW models was largely driven by their performance for small molecules on the S22 and S66 benchmark sets, and currently most popular DFT+vdW methods are able to achieve accuracies of 10–20 meV (better than 10%). The remaining errors are due to inaccuracy in the asymptotic vdW coefficients, empirical parameters in damping functions, and errors in the XC functional. Because of such rather uniform performance of different methods for small molecules, the focus has shifted to assessing the performance for larger systems. Here, in fact, the differences are more prominent, because the vdW energy makes a much larger relative contribution to cohesion. For example, for polarizable supramolecular systems, such as the "buckyball catcher" complex, pairwise dispersion corrections overestimate the binding energy by 0.4–0.6 eV, whereas including many-body dispersion effects reduces this error to 0.1 eV.

For periodic molecular crystals, some coarse-grained DFT+vdW methods are able to achieve remarkable accuracy of 40–50 meV per molecule (5%), compared to experimental results (Reilly and Tkatchenko 2013a, b; Brandenburg and Grimme 2014). Since the difference in lattice energies between various available experiments is on the same order of magnitude, this highlights the mature state of vdW dispersion corrections to semilocal DFT. The nonlocal density functionals yield somewhat larger errors (Otero-de-la-Roza and Johnson 2012). Understanding the performance of different vdW-inclusive methods for large molecular systems is still a subject of ongoing research. Accurate description of vdW interactions becomes even more relevant for the relative energetics of molecular systems, which are essential to predict the correct polymorphic behavior of molecular crystals (Reilly et al. 2016).

Beyond organic materials, the performance of DFT+vdW methods have started to be tested only recently. Reference lattice and interlayer binding energies calculated with RPA exist for a range of bulk solids (Harl et al. 2010; Schimka et al. 2013) and layered materials (Björkman et al. 2012; Björkman 2012, 2014), but the accuracy of RPA on the organic systems is comparable to that of DFT+vdW methods. Alternatively, experimental data are available for many of these systems. Of the different material types, the performance of individual methods for semiconductors is usually comparable to that of the organic compounds. In contrast, the accuracy with which ionic solids and metallic materials are described differs greatly between vdW models. In some cases, this has motivated the development of different flavors of vdW models, each targeting a specific class of materials.

## 7    Toward the Ultimate van der Waals Model

Despite the numerous advances in recent years discussed in the previous sections, a general, accurate, and efficient vdW model is not yet available. Arguably, this is a result of each of the current vdW models having some theoretical deficiencies, some of which are shared by all the models. Therefore, and to stimulate the reader with potential research problems, we conclude this chapter with a list of theoretical features that are missing from some or all current vdW models. Furthermore, we discuss which types of materials do we expect to benefit from potential addition of such features.

- All current vdW models assume a localized noninteracting polarizability, which is appropriate for gapped electronic systems, but not for gapless ones. At the same time, it is not clear what is the importance of the delocalized part of the response to vdW interactions in different circumstances. For instance, one can expect that the delocalized fluctuations will be manifested more strongly in the case of two metallic objects and less in the case of a molecule adsorbed on a metallic surface. Developing a unified model that uses both the localized and delocalized parts of the electronic response would enable treating the widest possible range of systems from purely covalent to purely metallic.

- Most current interatomic approaches assume a single-oscillator frequency dependence of the polarizability. This may be adequate for lighter elements of the periodic table in which majority of the response comes from the valence electrons with similar response properties. However, the core shells contribute nonnegligibly to the polarizability of heavier elements, and in such cases the use of two oscillators to capture separately the response of inner-shell and valence electrons could be beneficial to accuracy.
- All current models of effective polarizability assume isotropic local response. Could the directionality of the density gradient be used to construct an anisotropic polarizability model? The part of anisotropy resulting from the long-range interactions (for instance, due to specific packing in a molecular crystal) is already captured in many-body vdW effects. But the part stemming from short-range interactions between neighboring atoms (such as in planar aromatic compounds) is currently neglected.
- The VV functional is perhaps the most accurate semilocal functional for the local polarizability, but it is still lacking in accuracy to some of the interatomic approaches. Could this be potentially remedied by including dependence on the kinetic energy density or the second derivative of the electron density? Are there some exact constraints and limits on the polarizability functional? Improvements of this type can be expected to improve the accuracy of the vdW models across the whole spectrum of materials.
- All methods that use vdW models are based on some empirical coupling of the short-range and long-range parts. The empiricism necessarily hinders generality and introduces bias toward the system on which the coupling was constructed. Many theoretical results are available about general properties of XC functionals yet none so far that would enable a more rigorous coupling between semilocal DFT and vdW models. This question is especially relevant in systems where long-range vdW interactions contribute only partially to the total interaction energy, such as smaller organic complexes or organic/inorganic systems.
- Except for the costly RPA-based methods, there is no general vdW model that would be simultaneously many-body while treating also higher multipole moments of the polarizability. Some pairwise interatomic models include higher multipole moments, and nonlocal vdW functionals do not need to because of the lack of coarse-graining, but none of them are fully many-body methods. The MBD methods, on the other hand, have not yet been extended beyond the dipole approximation. The inclusion of higher multipoles would be beneficial for two reasons. First, it should improve the accuracy of the vdW model for strongly polarizable systems at equilibrium geometries. Second, it would enable to systematically study a whole new class of vdW effects such as the dependence of vdW interactions on external electric fields, which cannot be done within the dipole approximation.

# References

Ambrosetti A, Alfè D, DiStasio RA Jr, Tkatchenko A (2014a) Hard numbers for large molecules: toward exact energetics for supramolecular systems. J Phys Chem Lett 5(5):849–855. https://doi.org/10.1021/jz402663k

Ambrosetti A, Reilly AM, DiStasio RA Jr, Tkatchenko A (2014b) Long-range correlation energy calculated from coupled atomic response functions. J Chem Phys 140:18A508. https://doi.org/10.1063/1.4865104

Andersson Y, Langreth DC, Lundqvist BI (1996) Van der Waals interactions in density-functional theory. Phys Rev Lett 76(1):102–105. https://doi.org/10.1103/PhysRevLett.76.102

Ángyán JG (2007) On the exchange-hole model of London dispersion forces. J Chem Phys 127(2):024108. https://doi.org/10.1063/1.2749512

Ángyán JG, Liu RF, Toulouse J, Jansen G (2011) Correlation energy expressions from the adiabatic-connection fluctuation–dissipation theorem approach. J Chem Theory Comput 7(10):3116–3130. https://doi.org/10.1021/ct200501r

Axilrod BM, Teller E (1943) Interaction of the van der Waals type between three atoms. J Chem Phys 11(6):299–300. https://doi.org/10.1063/1.1723844

Ayers PW (2009) A perspective on the link between the exchange(-correlation) hole and dispersion forces. J Math Chem 46(1):86–96. https://doi.org/10.1007/s10910-008-9451-y

Bade WL (1957) Drude-model calculation of dispersion forces. I. General theory. J Chem Phys 27(6):1280–1284. https://doi.org/10.1063/1.1743991

Bade WL, Kirkwood JG (1957) Drude-model calculation of dispersion forces. II. The linear lattice. J Chem Phys 27(6):1284–1288. https://doi.org/10.1063/1.1743992

Becke AD (1988) A multicenter numerical integration scheme for polyatomic molecules. J Chem Phys 88(4):2547–2553. https://doi.org/10.1063/1.454033

Becke AD, Johnson ER (2005a) A density-functional model of the dispersion interaction. J Chem Phys 123(15):154101. https://doi.org/10.1063/1.2065267

Becke AD, Johnson ER (2005b) Exchange-hole dipole moment and the dispersion interaction. J Chem Phys 122(15):154104. https://doi.org/10.1063/1.1884601

Becke AD, Johnson ER (2006) Exchange-hole dipole moment and the dispersion interaction: high-order dispersion coefficients. J Chem Phys 124(1):014104. https://doi.org/10.1063/1.2139668

Becke AD, Roussel MR (1989) Exchange holes in inhomogeneous systems: a coordinate-space model. Phys Rev A 39(8):3761–3767. https://doi.org/10.1103/PhysRevA.39.3761

Björkman T (2012) Van der Waals density functional for solids. Phys Rev B 86(16):165109. https://doi.org/10.1103/PhysRevB.86.165109

Björkman T (2014) Testing several recent van der Waals density functionals for layered structures. J Chem Phys 141(7):074708. https://doi.org/10.1063/1.4893329

Björkman T, Gulans A, Krasheninnikov AV, Nieminen RM (2012) Are we van der Waals ready? J Phys Condens Matter 24(42):424218. https://doi.org/10.1088/0953-8984/24/42/424218

Brandenburg JG, Grimme S (2014) Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (DFTB). J Phys Chem Lett 5(11):1785–1789. https://doi.org/10.1021/jz500755u

Bučko T, Lebègue S, Hafner J, Ángyán JG (2013) Improved density dependent correction for the description of London dispersion forces. J Chem Theory Comput 9(10):4293–4299. https://doi.org/10.1021/ct400694h

Bučko T, Lebègue S, Ángyán JG, Hafner J (2014) Extending the applicability of the Tkatchenko-Scheffler dispersion correction via iterative Hirshfeld partitioning. J Chem Phys 141(3):034,114. https://doi.org/10.1063/1.4890003

Bultinck P, Alsenoy CV, Ayers PW, Carbó-Dorca R (2007) Critical analysis and extension of the Hirshfeld atoms in molecules. J Chem Phys 126(14):144111. https://doi.org/10.1063/1.2715563

Callen HB, Welton TA (1951) Irreversibility and generalized noise. Phys Rev 83(1):34–40. https://doi.org/10.1103/PhysRev.83.34

Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density functional for general geometries. Phys Rev Lett 92(24):246401. https://doi.org/10.1103/PhysRevLett.92.246401

Dobson JF, White A, Rubio A (2006) Asymptotics of the dispersion interaction: analytic benchmarks for van der Waals energy functionals. Phys Rev Lett 96(7):073,201. https://doi.org/10.1103/PhysRevLett.96.073201

Donchev AG (2006) Many-body effects of dispersion interaction. J Chem Phys 125(7):074713. https://doi.org/10.1063/1.2337283

Elstner M, Hobza P, Frauenheim T, Suhai S, Kaxiras E (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: a density-functional-theory based treatment. J Chem Phys 114(12):5149–5155. https://doi.org/10.1063/1.1329889

Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86(1):253–305. https://doi.org/10.1103/RevModPhys.86.253

Gao W, Tkatchenko A (2013) Electronic structure and van der Waals interactions in the stability and mobility of point defects in semiconductors. Phys Rev Lett 111(4):045501. https://doi.org/10.1103/PhysRevLett.111.045501

Ge X, Lu D (2015) Local representation of the electronic dielectric response function. Phys Rev B 92(24):241107. https://doi.org/10.1103/PhysRevB.92.241107

Geim AK, Grigorieva IV (2013) Van der Waals heterostructures. Nature 499(7459):419–425. https://doi.org/10.1038/nature12385

Gell-Mann M, Brueckner KA (1957) Correlation energy of an electron gas at high density. Phys Rev 106(2):364–368. https://doi.org/10.1103/PhysRev.106.364

Gobre VV, Tkatchenko A (2013) Scaling laws for van der Waals interactions in nanostructured materials. Nat Commun 4:2341. https://doi.org/10.1038/ncomms3341

Grimme S (2004) Accurate description of van der Waals complexes by density functional theory including empirical corrections. J Comput Chem 25(12):1463–1473. https://doi.org/10.1002/jcc.20078

Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem 27(15):1787–1799. https://doi.org/10.1002/jcc.20495

Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132(15):154104. https://doi.org/10.1063/1.3382344

Gunnarsson O, Lundqvist BI (1976) Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. Phys Rev B 13(10):4274–4298. https://doi.org/10.1103/PhysRevB.13.4274

Halgren TA (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. J Am Chem Soc 114(20):7827–7843. https://doi.org/10.1021/ja00046a032

Harl J, Schimka L, Kresse G (2010) Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids. Phys Rev B 81(11):115,126. https://doi.org/10.1103/PhysRevB.81.115126

Hepburn J, Scoles G, Penco R (1975) A simple but reliable method for the prediction of intermolecular potentials. Chem Phys Lett 36(4):451–456. https://doi.org/10.1016/0009-2614(75)80278-8

Hermann J, DiStasio RA Jr, Tkatchenko A (2017) First-principles models for van der Waals interactions in molecules and materials: concepts, theory, and applications. Chem Rev 117(6):4714–4758. https://doi.org/10.1021/acs.chemrev.6b00446

Heßelmann A (2009) Derivation of the dispersion energy as an explicit density- and exchange-hole functional. J Chem Phys 130(8):084104. https://doi.org/10.1063/1.3077939

Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theoret Chim Acta 44(2):129–138. https://doi.org/10.1007/BF00549096

Hunt KLC (1983) Nonlocal polarizability densities and van der Waals interactions. J Chem Phys 78(10):6149–6155. https://doi.org/10.1063/1.444577

Jansen G, Liu RF, Ángyán JG (2010) On the equivalence of ring-coupled cluster and adiabatic connection fluctuation-dissipation theorem random phase approximation correlation energy expressions. J Chem Phys 133(15):154106. https://doi.org/10.1063/1.3481575

Johnson ER, Becke AD (2006) A post-Hartree-Fock model of intermolecular interactions: inclusion of higher-order corrections. J Chem Phys 124(17):174104. https://doi.org/10.1063/1.2190220

Jurečka P, Šponer J, Černý J, Hobza P (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. Phys Chem Chem Phys 8(17):1985–1993. https://doi.org/10.1039/B600027D

Kohn W, Meir Y, Makarov DE (1998) Van der Waals energies in density functional theory. Phys Rev Lett 80(19):4153–4156. https://doi.org/10.1103/PhysRevLett.80.4153

Kurth S, Perdew JP (1999) Density-functional correction of random-phase-approximation correlation with results for jellium surface energies. Phys Rev B 59(16):10461–10468. https://doi.org/10.1103/PhysRevB.59.10461

Landau LD, Lifshitz EM (1980) Statistical physics, vol 5, 3rd edn. Pergamon Press, Oxford

Langbein D (1974) Theory of Van Der Waals attraction, Springer tracts in modern physics, vol 72. Springer, Berlin/Heidelberg. https://doi.org/10.1007/BFb0042407. http://link.springer.com/chapter/10.1007/BFb0042407

Langreth DC, Perdew JP (1977) Exchange-correlation energy of a metallic surface: wave-vector analysis. Phys Rev B 15(6):2884–2901. https://doi.org/10.1103/PhysRevB.15.2884

Lein M, Gross EKU, Perdew JP (2000) Electron correlation energies from scaled exchange-correlation kernels: importance of spatial versus temporal nonlocality. Phys Rev B 61(20):13431–13437. https://doi.org/10.1103/PhysRevB.61.13431

Lucas A (1967) Collective contributions to the long-range dipolar interaction in rare-gas crystals. Physica 35(3):353–368. https://doi.org/10.1016/0031-8914(67)90184-X

Mahan GD (1965) Van der Waals forces in solids. J Chem Phys 43(5):1569–1574. https://doi.org/10.1063/1.1696973

Margenau H (1939) Van der Waals forces. Rev Mod Phys 11(1):1–35. https://doi.org/10.1103/RevModPhys.11.1

Marzari N, Mostofi AA, Yates JR, Souza I, Vanderbilt D (2012) Maximally localized Wannier functions: theory and applications. Rev Mod Phys 84(4):1419–1475. https://doi.org/10.1103/RevModPhys.84.1419

Mayer A (2007) Formulation in terms of normalized propagators of a charge-dipole model enabling the calculation of the polarization properties of fullerenes and carbon nanotubes. Phys Rev B 75(4):045407. https://doi.org/10.1103/PhysRevB.75.045407

Mooij WTM, van Duijneveldt FB, van Duijneveldt-van de Rijdt JGCM, van Eijck BP (1999) Transferable ab initio intermolecular potentials. 1. Derivation from methanol dimer and trimer calculations. J Phys Chem A 103(48):9872–9882. https://doi.org/10.1021/jp991641n

Mutō Y (1943) Hikyokusei bunshi no aida ni sayō suru chikara ni tsuite [On forces acting between nonpolar molecules]. Nippon Sugaku Buturigakkwaishi 17:629–631. https://doi.org/10.11429/subutsukaishi1927.17.10-11-12_629

Olsen T, Thygesen KS (2012) Extending the random-phase approximation for electronic correlation energies: the renormalized adiabatic local density approximation. Phys Rev B 86(8):081103. https://doi.org/10.1103/PhysRevB.86.081103

Olsen T, Thygesen KS (2013a) Beyond the random phase approximation: improved description of short-range correlation by a renormalized adiabatic local density approximation. Phys Rev B 88(11):115131. https://doi.org/10.1103/PhysRevB.88.115131

Olsen T, Thygesen KS (2013b) Random phase approximation applied to solids, molecules, and graphene-metal interfaces: from van der Waals to covalent bonding. Phys Rev B 87(7):075111. https://doi.org/10.1103/PhysRevB.87.075111

Olsen T, Thygesen KS (2014) Accurate ground-state energies of solids and molecules from time-dependent density-functional theory. Phys Rev Lett 112(20):203001. https://doi.org/10.1103/PhysRevLett.112.203001

Otero-de-la-Roza A, Johnson ER (2012) A benchmark for non-covalent interactions in solids. J Chem Phys 137(5):054103. https://doi.org/10.1063/1.4738961

Parr RG, Yang W (1989) Density-functional theory of atoms and molecules. No. 16 in The international series of monographs on chemistry. Oxford University Press. https://global.oup.com/academic/product/density-functional-theory-of-atoms-and-molecules-9780195092769

Parsegian VA (2005) Van der Waals forces: a handbook for biologists, chemists, engineers, and physicists. Cambridge University Press, Cambridge. http://ebooks.cambridge.org/ref/id/CBO9780511614606

Reilly AM, Tkatchenko A (2013a) Seamless and accurate modeling of organic molecular materials. J Phys Chem Lett 4(6):1028–1033. https://doi.org/10.1021/jz400226x

Reilly AM, Tkatchenko A (2013b) Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. J Chem Phys 139(2):024,705. https://doi.org/10.1063/1.4812819

Reilly AM, Cooper RI, Adjiman CS, Bhattacharya S, Boese AD, Brandenburg JG, Bygrave PJ, Bylsma R, Campbell JE, Car R, Case DH, Chadha R, Cole JC, Cosburn K, Cuppen HM, Curtis F, Day GM, DiStasio RA Jr, Dzyabchenko A, van Eijck BP, Elking DM, van den Ende JA, Facelli JC, Ferraro MB, Fusti-Molnar L, Gatsiou CA, Gee TS, de Gelder R, Ghiringhelli LM, Goto H, Grimme S, Guo R, Hofmann DWM, Hoja J, Hylton RK, Iuzzolino L, Jankiewicz W, de Jong DT, Kendrick J, de Klerk NJJ, Ko HY, Kuleshova LN, Li X, Lohani S, Leusen FJJ, Lund AM, Lv J, Ma Y, Marom N, Masunov AE, McCabe P, McMahon DP, Meekes H, Metz MP, Misquitta AJ, Mohamed S, Monserrat B, Needs RJ, Neumann MA, Nyman J, Obata S, Oberhofer H, Oganov AR, Orendt AM, Pagola GI, Pantelides CC, Pickard CJ, Podeszwa R, Price LS, Price SL, Pulido A, Read MG, Reuter K, Schneider E, Schober C, Shields GP, Singh P, Sugden IJ, Szalewicz K, Taylor CR, Tkatchenko A, Tuckerman ME, Vacarro F, Vasileiadis M, Vazquez-Mayagoitia A, Vogt L, Wang Y, Watson RE, de Wijs GA, Yang J, Zhu Q, Groom CR (2016) Report on the sixth blind test of organic crystal structure prediction methods. Acta Cryst B 72(4):439–459. https://doi.org/10.1107/S2052520616007447

Ren X, Tkatchenko A, Rinke P, Scheffler M (2011) Beyond the random-phase approximation for the electron correlation energy: the importance of single excitations. Phys Rev Lett 106(15):153003. https://doi.org/10.1103/PhysRevLett.106.153003

Ren X, Rinke P, Joas C, Scheffler M (2012) Random-phase approximation and its applications in computational chemistry and materials science. J Mater Sci 47(21):7447–7471. https://doi.org/10.1007/s10853-012-6570-4

Ren X, Rinke P, Scuseria GE, Scheffler M (2013) Renormalized second-order perturbation theory for the electron correlation energy: concept, implementation, and benchmarks. Phys Rev B 88(3):035120. https://doi.org/10.1103/PhysRevB.88.035120

Renne MJ, Nijboer BRA (1967) Microscopic derivation of macroscopic Van der Waals forces. Chem Phys Lett 1(8):317–320. https://doi.org/10.1016/0009-2614(67)80004-6

Řezáč J, Riley KE, Hobza P (2011) S66: a well-balanced database of benchmark interaction energies relevant to biomolecular structures. J Chem Theory Comput 7(8):2427–2438. https://doi.org/10.1021/ct2002946

Ruiz VG, Liu W, Zojer E, Scheffler M, Tkatchenko A (2012) Density-functional theory with screened van der Waals interactions for the modeling of hybrid inorganic-organic systems. Phys Rev Lett 108(14):146103. https://doi.org/10.1103/PhysRevLett.108.146103

Ruzsinszky A, Perdew JP, Tao J, Csonka GI, Pitarke JM (2012) Van der Waals coefficients for nanostructures: fullerenes defy conventional wisdom. Phys Rev Lett 109(23):233203. https://doi.org/10.1103/PhysRevLett.109.233203

Sato T, Nakai H (2009) Density functional method including weak interactions: dispersion coefficients based on the local response approximation. J Chem Phys 131(22):224104. https://doi.org/10.1063/1.3269802

Sato T, Nakai H (2010) Local response dispersion method. II. Generalized multicenter interactions. J Chem Phys 133(19):194101. https://doi.org/10.1063/1.3503040

Schimka L, Gaudoin R, Klimeš J, Marsman M, Kresse G (2013) Lattice constants and cohesive energies of alkali, alkaline-earth, and transition metals: random phase approximation and den-

sity functional theory results. Phys Rev B 87(21):214102. https://doi.org/10.1103/PhysRevB.87.214102

Scuseria GE, Henderson TM, Sorensen DC (2008) The ground state correlation energy of the random phase approximation from a ring coupled cluster doubles approach. J Chem Phys 129(23):231101. https://doi.org/10.1063/1.3043729

Silvestrelli PL (2008) Van der Waals interactions in DFT made easy by Wannier functions. Phys Rev Lett 100(5):053002. https://doi.org/10.1103/PhysRevLett.100.053002

Silvestrelli PL (2013) Van der Waals interactions in density functional theory by combining the quantum harmonic oscillator-model with localized Wannier functions. J Chem Phys 139(5):054,106. https://doi.org/10.1063/1.4816964

Silvestrelli PL, Benyahia K, Grubisiĉ S, Ancilotto F, Toigo F (2009) Van der Waals interactions at surfaces by density functional theory using Wannier functions. J Chem Phys 130(7):074,702. https://doi.org/10.1063/1.3077288

Steinmann SN, Corminboeuf C (2010) A system-dependent density-based dispersion correction. J Chem Theory Comput 6(7):1990–2001. https://doi.org/10.1021/ct1001494

Steinmann SN, Corminboeuf C (2011) Comprehensive benchmarking of a density-dependent dispersion correction. J Chem Theory Comput 7(11):3567–3577. https://doi.org/10.1021/ct200602x

Stone AJ (2013) The theory of intermolecular forces, 2nd edn. Oxford University Press, Oxford http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199672394.001.0001/acprof-9780199672394

Szabo A, Ostlund NS (1996) Modern quantum chemistry: introduction to advanced electronic structure theory. McGraw-Hill, New York (reprint). http://store.doverpublications.com/0486691861.html

Tao J, Perdew JP, Ruzsinszky A (2010) Long-range van der Waals attraction and alkali-metal lattice constants. Phys Rev B 81(23):233,102. https://doi.org/10.1103/PhysRevB.81.233102

Tkatchenko A, Scheffler M (2009) Accurate molecular Van Der Waals interactions from ground-state electron density and free-atom reference data. Phys Rev Lett 102(7):073005. https://doi.org/10.1103/PhysRevLett.102.073005

Tkatchenko A, DiStasio RA Jr, Car R, Scheffler M (2012) Accurate and efficient method for many-body van der Waals interactions. Phys Rev Lett 108(23):236402. https://doi.org/10.1103/PhysRevLett.108.236402

Tkatchenko A, Ambrosetti A, DiStasio RA Jr (2013) Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. J Chem Phys 138(7):74106. https://doi.org/10.1063/1.4789814

Toulouse J, Colonna F, Savin A (2004) Long-range–short-range separation of the electron-electron interaction in density-functional theory. Phys Rev A 70(6):062505. https://doi.org/10.1103/PhysRevA.70.062505

Toulouse J, Gerber IC, Jansen G, Savin A, Ángyán JG (2009) Adiabatic-connection fluctuation-dissipation density-functional theory based on range separation. Phys Rev Lett 102(9):096404. https://doi.org/10.1103/PhysRevLett.102.096404

Unsöld A (1927) Quantentheorie des Wasserstoffmolekülions und der Born-Landéschen Abstoßungskräfte. Z Physik 43(8):563–574. https://doi.org/10.1007/BF01397633

Verstraelen T, Pauwels E, De Proft F, Van Speybroeck V, Geerlings P, Waroquier M (2012) Assessment of atomic charge models for gas-phase computations on polypeptides. J Chem Theory Comput 8(2):661–676. https://doi.org/10.1021/ct200512e

Vydrov OA, Van Voorhis T (2010) Nonlocal van der Waals density functional: the simpler the better. J Chem Phys 133(24):244103. https://doi.org/10.1063/1.3521275

Wu Q, Yang W (2002) Empirical correction to density functional theory for van der Waals interactions. J Chem Phys 116(2):515–524. https://doi.org/10.1063/1.1424928

Zhang GX, Tkatchenko A, Paier J, Appel H, Scheffler M (2011) Van der Waals interactions in ionic and semiconductor solids. Phys Rev Lett 107(24):245501. https://doi.org/10.1103/PhysRevLett.107.245501

# Pump-Probe Photoelectron Spectra

# 14

Umberto De Giovannini

## Contents

**Abstract**

Pump-probe photoelectron spectroscopy provides a tool to observe excitations taking place in electronic systems as they evolve in time. This technique is frequently applied to study complex phenomena taking place in chemistry and solid-state physics. To properly capture the dynamics observed in the experiments, one needs to employ non-perturbative theories capable to describe the complete time evolution of large physical systems. After a pedagogical survey on the literature, in this chapter, we focus on TDDFT and illustrate how this theory can be formulated in a way that can capture the entire ionization dynamics in atoms, molecules, and solids.

## 1    Overview

Photoelectron spectroscopy (PES) is nowadays one the most prominent experimental tools employed to measure the electronic structure of matter in the solid, gas,

U. De Giovannini (✉)

Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany
e-mail: umberto.degiovannini@gmail.com

and liquid phases. In this chapter we are concerned with the application of PES to two classes of electronic systems: *finite* systems, such as atoms and molecules in gas phase, and *infinitely* periodic systems, such as bulk crystals and surfaces. This spectroscopy is based on the photoelectric effect, the physical phenomenon where electrons are extracted from a sample by illuminating it with light of an appropriate wavelength which was first discovered by Hertz in 1887. Electrons emerging from this kind of process are normally called photoelectrons.

The fundamental equation governing photo-ionization is the Einstein law which states the energy balance between the kinetic energy of the ionized electrons $E_{kin}$, the energy of the photon $\hbar\omega$, and the ionization energy, or ionization potential, $I_p$, that keeps them bind to the sample. Depending on the context, this relation is expressed in slightly different ways. In atomic and molecular physics, it usually reads: $E_{kin} = \hbar\omega - I_p$. In solid-state and surface physics, it is customary to break down the ionization energy as the sum of the work function $\phi$ and the binding energy $E_b$ both measured from the Fermi energy as illustrated in Fig. 1. In this case the relation becomes $E_{kin} = \hbar\omega - \phi - E_b$. Obviously, the underlying mechanism remains the same.

From these basic considerations, the spectroscopic potential of the photoelectron observables is immediately apparent because, by measuring the kinetic energy of the ionized electrons and knowing the frequency of the field, one can obtain a picture of the electronic energy levels of the system. It must be noted, however, that this technique can only access the properties of occupied states. To measure empty states, one in principle can use the inverse process where an electron is captured by the material and emits a photon.

In the literature, it is customary to divide photoelectron spectroscopy in two main categories depending on the radiation used. We are dealing with ultraviolet PES, if the field is in the UV range and only ionizes valence electrons, while we have X-ray PES if the field is energetic enough to ionize inner shell electrons. However, based on the same physical process, there are theoretical details specific for each energy
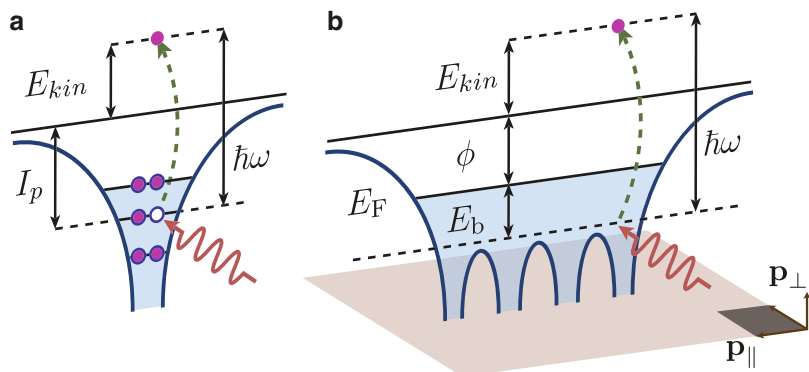


**Fig. 1** Schematic of photoelectron spectroscopy in the case of (**a**) finite systems, such as atoms and molecules, and (**b**) infinitely periodic systems, i.e., bulk crystals

scale, most notably whether or not it is allowed to invoke the dipole approximation. The discussion of this chapter is concerned with photon energies up to the XUVs ($\approx 100$ eV) where the dipole approximation is well justified.

The intensity of the photoelectron current, or equivalently the probability rate to eject photoelectrons with a given velocity, is the observable quantity in the experiments. It can be estimated using Fermi's golden rule

$$I = \frac{2\pi}{\hbar} \sum_f \left| \left\langle f \left| \frac{e}{mc} \mathbf{A} \cdot \hat{\mathbf{p}} \right| i \right\rangle \right|^2 \delta(E_f - E_i - \hbar\omega). \tag{1}$$

In the equation, $|i\rangle$ is the many-body $N$-electrons ground state representing the system in its initial state, while $|f\rangle$ is the state describing a single ionization and is composed of $N - 1$ bound electrons plus one in the continuum with momentum $\mathbf{p}$.

By summing overall the possible final states that satisfy the energy conservation – the Einstein relation – imposed by the delta function, one recovers the full distribution in momentum, $I(\mathbf{p})$. This quantity goes under different names depending on the community of reference. In atomic and molecular physics, it is oftentimes called photoelectron angular distribution (PAD), while in the solid-state community, it goes under the name of angular-resolved photoelectron spectroscopy (ARPES) spectrum. The reasons for this apparent dichotomy originate from the differences in the experimental detectors and the specific spectroscopic information attainable in the different contexts. In solid-state physics, the usual assumption is that, during ionization, the photoelectron's momentum parallel to the surface is conserved and equal to the crystal momentum of the electrons in the sample, $\mathbf{p}_{\parallel} = \mathbf{k}_{\parallel}$. Therefore the intensity expressed as a function of the binding energy and the parallel momentum, $I(\mathbf{p}_{\parallel}, E_{\text{kin}})$, measures the dispersion of the energy levels of the system, i.e., the band structure. For molecules the physical content of $I(\mathbf{p})$ is still in the process of beginning to be fully understood, for instance, in planar molecules, it was shown to be linked with the molecular orbital spatial distribution (Puschnig et al. 2009).

The picture given by (1) is an idealization of the photoemission process that does not take into account electron-electron interaction and surface effects. In fact, the electrons in the system have to relax in order to screen the hole left behind after ionization, and, during this dynamical process, they can excite other modes of the sample. In solids, for instance, it is often possible to observe charge density oscillations – plasmons – or lattice vibrations, phonons, being excited by the ejected electrons on their way to the vacuum. These processes ultimately affect the kinetic energy distribution of the photoelectrons by introducing new peaks and by imposing a finite peak width that reflects the lifetime of the excitation. To include these effects in (1), the delta function has to be substituted by the more general spectral function $A(\omega, \mathbf{p})$. The spectral function is a many-body object which can be rigorously defined in terms of the one-particle Green's function, and it is usually calculated in the so-called *GW approximation*. For more details, we remind to the review article (Damascelli et al. 2003) and references therein.

After substituting the delta with the spectral function, to get the full ionization current from (1), one has to evaluate the dipole matrix elements. These terms contain the photoelectron current dependence on the external field properties such as photon polarization and energy and are crucial for the complete account of the experimental spectrum. To calculate the matrix elements, one needs to explicitly evaluate the continuum states of the Hamiltonian that constitutes the scattering-wave component – i.e., the ionized electron – of the final state $|f\rangle$. The simplest approximation often employed in the literature is to substitute the scattering wave with a plane wave $|\mathbf{p}\rangle$, so that the matrix element turns into a simple Fourier transform of the initial wavefunction: for one-electron systems, $\langle\mathbf{p}|\mathbf{A}\cdot\hat{\mathbf{p}}|i\rangle = \mathbf{A}\cdot\mathbf{p}\langle\mathbf{p}|i\rangle$. The quality of this approximation varies from system to system and is not always good especially for low-energy continuum states that are strongly affected by the electrostatic potential. For a better agreement to experiments, one has to turn to more sophisticated approaches. The time-reversed low-energy electron-diffraction – known as LEED – states (Pendry 1990) in solid-state physics and the B-spline basis (Bachau et al. 2001) in atomic and molecular physics are among the most popular expansions.

Non-perturbative approaches alternative to (1), like the direct solution of the time-dependent Schrödinger equation (TDSE), can also be used to model the full ionization process. This is a usual practice in the atomic strong-field community where the fields are so strong to allow multiphoton processes to take place and where the interference pattern formed by the photoelectrons ejected at different times in the pulse encodes a plethora of valuable information. In these cases, the photoelectron spectrum can be directly obtained by projecting the long time limit of the time-dependent wavefunction $|\Psi(t)\rangle$ on the continuum states: $I \propto \langle f|\Psi(t)\rangle$. Even in this approach, a detailed account of the continuum state is therefore needed, which oftentimes is already present in the basis used to integrate the TDSE.

Motivated by the need of a non-perturbative theory, the strong-field community developed methods to calculate the photoelectron probability that do not require any explicit knowledge of the continuum states; the resolvent technique (Catoire and Bachau 2012) and the t-SURFF method (Tao and Scrinzi 2012) are the most notable results of this effort. These methods have been successfully employed to calculate PES and PADs in atoms under strong and long wavelength pulses where electrons follow large rescattering trajectories before being released into the continuum (Blaga et al. 2009). Clearly, the application of a non-perturbative approach to the single photon ionization described so far is straightforward, even though the solution of the TDSE may not be very efficient or properly motivated. When PES is employed to study systems out of equilibrium, however, the non-perturbative approach becomes more necessary.

Time-resolved photoelectron spectroscopy extends PES to excited states in the time domain. In this extension the spectra are usually acquired in pump-probe setups where two laser pulses are directed onto the sample and where one of the pulses, the pump, creates an excitation in the system which is then observed by ionizing it with the probe. The presence of two pulses breaks the time translation symmetry, and therefore the spectra acquire an explicit dependence on the relative delay between the pulses. Recording different spectra as a function of this delay constitutes the

time-resolved spectrum from which one can access information about the dynamics of the excitation.

In pump-probe experiments, by selecting the length of the pulses, one can observe the time trace of different physical excitations. On the femtosecond scale, the scale of ionic motion, time-resolved measurements have been applied to study chemical reactions ranging from bond breaking in dimers to nuclear motion in large biological molecules. This line of research developed into the field of *femtochemistry* (Stolow et al. 2004) which culminated with the Nobel Prize in Chemistry to A. H. Zewail in 1999. With shorter pulses, one can further increase time resolution and observe even faster dynamics. The shortest time scale attainable today is the attosecond scale, the characteristic scale of the electronic excitations. It lays at the forefront of the research as it combines complex experiments with the theoretical challenges presented by the description of coherent electron dynamics (Lépine et al. 2014). In solid-state physics, the use of ultrafast time-resolved approaches started later than molecular physics but quickly caught up, and the last decade has seen these techniques applied to study many different processes such as, for instance, the exciton relaxation and excitation dynamics or the ultrafast charge migration and screening in insulators (Krausz and Stockman 2014).

Given the dynamical nature of the physical observables accessible via pump-probe PES, it is clear that any theoretical approach used to interpret the results of the experiments must be capable to account for the time evolution of the system. In this respect, we must first distinguish between two substantially different regimes depending on whether the pulses overlap or not. When the pulses overlap, the system is governed by a time-dependent Hamiltonian which explicitly contains the pump field and that can be quite different from the one of the samples in equilibrium. This is no longer the case when the two pulses do not overlap because the Hamiltonian is the same as the one in equilibrium.

In the nonoverlapping regime, the system finds itself into a state which, in the most general case, is in a superposition of different excited states of the equilibrium Hamiltonian. In this case one can still use (1) by choosing as initial state $|i\rangle$ the pump-generated wavepacket, $|\Psi(t)\rangle$, which is expanded in the basis of the exact excited states $|\Psi_i\rangle$ with eigenvalues $E_i$: $|\Psi(t)\rangle = \sum_i A_i e^{-i(E_i t/\hbar + \phi_i)}|\Psi_i\rangle$. Here $\phi_i$ and $A_i$ are, respectively, the phases and expansion coefficients of the wavepacket and are completely determined by the pump. The strength of this formulation is that, for isolated systems, the coefficients are time independent and therefore the time propagation can be carried out analytically. In practice, since the exact eigenstates are not easily available, one usually has to resort to approximate eigenstates of a simplified Hamiltonian, and therefore the coefficients become time dependent (Wu et al. 2011). Besides the specific approximations concerning the details of the physical system under study, one limitation of this approach is that the construction of the initial wavepacket can be accurately approximated only for a limited number of cases. To cover the most general situation, one needs to simulate the pump excitation process with a non-perturbative approach.

Non-equilibrium Green's function theory offers, in principle, a more general platform to formulate the out-of-equilibrium many-body problem (Kadanoff and

Baym 1962). In this context the most promising approach to date is based on the Kadanoff-Baym formulation for the lesser Green's function $G^<$ (Perfetto et al. 2016a). Given the computational cost of the implementation, this approach has been so far limited to simple systems, e.g., Perfetto et al. (2016b).

When pump and probe overlap, PES carries spectroscopic information about the pump-dressed system. The interest in this regime is that, since the pump field directly enters the Hamiltonian, it can give a handle to parametrically modify it. This modification can possibly result in new physical properties of the driven system that can be observed by PES (Hsieh et al. 2017). From the basic standpoint, the properties of the periodically driven system can be understood in the framework of Floquet theory. However, to properly account for the PES even in this regime, it appears necessary to simulate the full time-dependent evolution.

After this brief and non-exhaustive survey of the theoretical approaches available to model pump-probe PES, it emerges that, in order to include the largest possible spectrum of experimental conditions, one needs to simulate the full photo-ionization process as it develops in time. Our choice for this chapter is time-dependent density functional theory (TDDFT). TDDFT is an exact reformulation of the TDSE in terms of the time-dependent density which has been successfully used to model the dynamics of atomic, molecular, and solid-state systems. This flexibility offers the opportunity to illustrate the application of the theory to finite systems and infinitely periodic systems on equal footing.

In the remainder of this chapter, unless otherwise specified, we use Hartree atomic units where $\hbar = e = m_e = 4\pi\varepsilon_0 = 1$.

## 2    Real-Time TDDFT

Below, we briefly recall the relevant concepts of TDDFT, already introduced in the plenary ▶ Chap. 6 ,"TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter," and the specific features of the real-time approach that are instrumental for the description of pump-probe photoemission.

The central idea of TDDFT is to express the equation of motion for the time-dependent density, $n(\mathbf{r}, t)$, of an interacting many-body system as a set of equations for the density of a fictitious auxiliary system of noninteracting electrons – the *Kohn-Sham* (KS) system. The resulting equations – the time-dependent KS (TDKS) equations – are expressed in terms of the orbitals $\varphi_i$ of a single KS Slater determinant as follows:

$$i\frac{\partial}{\partial t}\varphi_i(\mathbf{r}, t) = H_{\text{KS}}(\mathbf{r}, t)\varphi_i(\mathbf{r}, t), \quad i = 1, \ldots, N/2, \tag{2}$$

$$H_{\text{KS}}(\mathbf{r}, t) = \frac{1}{2}\left(-i\nabla - \frac{\mathbf{A}(t)}{c}\right)^2 + v_{\text{KS}}[n](\mathbf{r}, t)$$

$$n(\mathbf{r}, t) = 2\sum_{i=1}^{N/2}|\varphi_i(\mathbf{r}, t)|^2. \tag{3}$$

To avoid nonessential complications in the notation, here we choose to describe systems with an even number of electrons $N$, so that each spatial orbital $\varphi_i$ is doubly occupied with two electrons of opposite spin. The time-dependent density $n(\mathbf{r}, t)$ corresponds both to the real and to the KS system, thanks to the action of the KS potential $v_{KS}$, which is composed of

$$v_{KS}[n](\mathbf{r}, t) = v_{ion}(\mathbf{r}, t) + v_H[n](\mathbf{r}, t) + v_{xc}[n](\mathbf{r}, t). \tag{4}$$

In the definition, the first term, $v_{ion}(\mathbf{r}, t)$, is the electron-ion potential provided by the nuclei, while the second term is the classical electrostatic, or Hartree, potential $v_H[n](\mathbf{r}, t)$, and the last term $v_{xc}[n](\mathbf{r}, t)$ is the so-called exchange and correlation potential, whose form is unknown and must be approximated.

For observables involving ionization, like PES, it is crucial to choose a functional capable to provide a good ionization energy. Exception made from hybrid functionals, containing the computationally expensive *exact-exchange* functional – i.e., the Fock operator – the large majority of functionals is unable to cancel the self-interaction component of the Hartree potential, self-interaction error, which results in an unphysical asymptotic exponential decay instead of the correct $\propto 1/r$ given by the electrostatic potential of a point charge. Overall this results in a systematic underestimation of the ionization energy. Many self-interaction correction (SIC) strategies have been developed over the years, but, owing to its simplicity, the average-density SIC (Legrand et al. 2002) has been so far the most adopted.

Another important aspect that has to be considered in order to properly describe ionization is charge removal. Any numerical implementations aiming at solving the TDKS equations have to deal with a finite basis set or simulation box which is incompatible with a correct description of ionization where, by definition, electrons are removed from a system and taken to infinity. To simulate the presence of an infinite surrounding vacuum space, it is customary to use absorbing boundaries. These boundaries act to prevent spurious reflections when the electronic density reaches the limit of the numerical representation mimicking the behavior of an open system. Note that, in the presence of these boundaries, the total charge is not conserved and the Hamiltonian is no longer Hermitian. The choice is between complex absorbing potentials and exterior complex scaling. While this last one is in principle superior, its numerical implementation is quite involved, and in practice very few codes use it. Complex potentials, though much easier to implement, pay the price of being not perfectly absorbing. This translates into having to deal with a window of kinetic energies for which the transparency is achieved that depends on the functional form of the potential and that has to be carefully tuned on the basis of the expected ionization dynamics (De Giovannini et al. 2015).

To simulate pump-probe spectra, one needs to couple the system with an external laser field. In (2), this coupling is expressed in the so-called *velocity gauge* with the field represented by a classical time-dependent vector potential $\mathbf{A}(t)$ in the dipole approximation. This means that the vector potential is taken to be constant in space, a good approximation as long as the field's wavelength is much larger than the spatial extension of the system – up to soft X-rays for systems smaller than 1 nm.

The use of the velocity gauge is uncommon in atomic and molecular physics where the coupling is usually expressed in terms of the electric field, $E \cdot \mathbf{r}$, in the so-called *length gauge*. The reason for this choice is functional to our scope. That is because the velocity gauge is compatible with both finite and periodic systems and, as we will show in the next section, it is particularly convenient to describe the coupling with ionized free electrons.

The choice of describing the coupling with a classical field instead of a quantized one is well justified in those cases, like in modern pump-probe experiments, where the laser fields employed have large fluences and are thus constituted by a large number of photons. In particular since $\mathbf{A}(t)$ is a general function of time, it is well suited to describe any linear combination of laser pulses and thus can perfectly account for any pump-probe configuration. It is important to stress that, given the nature of the pulses, it is not possible to invoke any linear response approach and the TDKS equations have to be integrated by real-time propagation. A large array of propagation options is available in the literature to perform this task (Castro et al. 2004).

The TDKS equations are general to any atomic and molecular system regardless of their shape and size and, in principle, are suited to describe systems, such as bulk crystals, which are microscopically infinite. Considering an infinitely extended $\varphi_i(\mathbf{r}, t)$ is of course not practical, but, exploiting the translation symmetry, it is possible to substantially simplify the formulation of the problem. Since the crystal periodically repeats in space, we can describe the infinite lattice with wavefunctions defined in a primitive cell, and using the Bloch theorem, we can express the KS orbitals as Bloch waves

$$\varphi_{i,\mathbf{k}}(\mathbf{r}, t) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{i,\mathbf{k}}(\mathbf{r}, t) \tag{5}$$

where $u_{i,\mathbf{k}}(\mathbf{r}, t)$ is a wavefunction with the periodicity of the lattice and $\mathbf{k}$ spans the first Brillouin zone (BZ) in reciprocal space.

By plugging the expression for the Bloch wave into (2), we arrive to a new set of TDKS equations only for the periodic component of the KS orbitals

$$i\frac{\partial}{\partial t} u_{i,\mathbf{k}}(\mathbf{r}, t) = \frac{1}{2}\left(-i\nabla + \mathbf{k} - \frac{\mathbf{A}(t)}{c}\right)^2 u_{i,\mathbf{k}}(\mathbf{r}, t) + v_{\mathrm{KS}}[n](\mathbf{r}, t) u_{i,\mathbf{k}}(\mathbf{r}, t),$$
$$i = 1, \ldots, N/2, \tag{6}$$

$$n(\mathbf{r}, t) = 2\sum_{i=1}^{N/2}\int_{BZ} d\mathbf{k} \, |\varphi_{i,\mathbf{k}}(\mathbf{r}, t)|^2 = 2\sum_{i=1}^{N/2}\int_{BZ} d\mathbf{k} \, |u_{i,\mathbf{k}}(\mathbf{r}, t)|^2. \tag{7}$$

These equations govern the time evolution of a generic nonmagnetic bulk insulator where each of the $N/2$ lowest-lying orbital states accommodates two electrons of opposite spin.
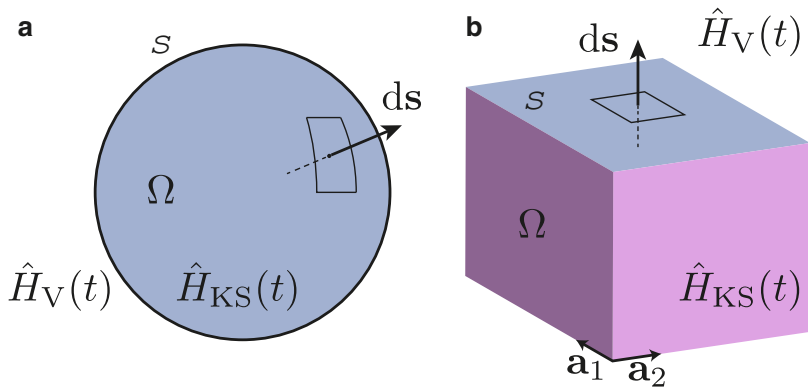
**Fig. 2** Scheme illustrating the geometries employed to calculate the photoelectron spectra for (**a**) finite and (**b**) infinitely periodic systems

## 3 Photoemission from Atoms and Molecules

The outcome of a ionization process is characterized by the total ionization probability $\mathscr{P}$: the probability to find a system, an atom, or a molecule, in a determined charged state after ionization. Given a spherical region $\Omega$ surrounding the system, like in Fig 2a, one can calculate $\mathscr{P}$ by integrating the charge density in its complement $\bar{\Omega}$, which is equivalent to the number of escaped electrons $N_{esc}(t)$, any time $t > t'$ after the ionization has taken place. In the long time limit, it follows from this definition that

$$\mathscr{P} = \lim_{t \to \infty} \frac{N_{esc}(t)}{N} = \lim_{t \to \infty} \frac{\int_{\bar{\Omega}} d\mathbf{r} \, n(\mathbf{r}, t)}{N}. \tag{8}$$

As we discussed at the beginning of this chapter, many important physical properties can be obtained only by looking at more *resolved* ionization observables than just the total probability. These include the energy-resolved or momentum-resolved photoelectron probabilities, $\mathscr{P}(E)$ and $\mathscr{P}(\mathbf{p})$, which can be obtained from $\mathscr{P}$ by direct differentiation

$$\mathscr{P}(E) = \frac{\partial \mathscr{P}}{\partial E} \text{ and } \mathscr{P}(\mathbf{p}) = \frac{\partial \mathscr{P}}{\partial \mathbf{p}}. \tag{9}$$

In order to use this definition, one needs to *expose* the functional dependence of $\mathscr{P}$ from the differentiating variable, $\mathbf{p}$ or $E$, and in what follows, we employ the flux of the ionization current through a closed surface to accomplish this task. This approach is based on the t-SURFF method, first introduced by Scrinzi (Tao and Scrinzi 2012) in the strong-field community for single-electron systems. Since here we are interested in many-electron systems described with TDDFT, we present the extended version of this method (Wopperer et al. 2017).

Note that the two quantities in (9) are not independent and one can recover one from the other using $E = \mathbf{p}^2/2$ – the free electron energy dispersion. In particular $\mathscr{P}(\mathbf{p})$, being resolved in momentum is the most general one. For this reason we can concentrate only on $\mathscr{P}(\mathbf{p})$, granted the other quantities can be derived from it by integration.

First we postulate that the KS Hamiltonian $\hat{H}(\mathbf{r}, t)$ describing a photo-ionization process can be factored out into two spatially separated components, one describing fully interacting electrons in an area surrounding the system, $\Omega$, and another one representing the asymptotic behavior of quasi-free electrons in the ionization region, $\bar{\Omega}$. In practice this is done assuming that the TDKS Hamiltonian (2) asymptotically reduces to a Volkov Hamiltonian in $\bar{\Omega}$:

$$
\hat{H}(\mathbf{r}, t) = \begin{cases} \hat{H}_{\mathrm{KS}}(\mathbf{r}, t) \text{ if } \mathbf{r} \in \Omega \\ \hat{H}_{\mathrm{V}}(\mathbf{r}, t) \text{ if } \mathbf{r} \in \bar{\Omega} \end{cases}. \tag{10}
$$

The Volkov Hamiltonian describes noninteracting free electrons driven by an external vector potential $\mathbf{A(t)}$ and is defined as

$$
\hat{H}_{\mathrm{V}}(\mathbf{r}, t) = \frac{1}{2}\left(-i\nabla - \frac{\mathbf{A(t)}}{c}\right)^2. \tag{11}
$$

Note that the partitioning (10) is an idealization since we are discarding the long-range tails of the Coulomb potential always present in the real systems. We are thus committing an error that is proportional to the radius of $\Omega$ and (10) that has to be considered as an approximation. The price that we pay is well compensated by the properties of the Volkov Hamiltonian. In particular the fact that it is analytically solvable and its time-dependent solutions – the Volkov waves – defined by

$$
\chi_{\mathbf{p}}(\mathbf{r}, t) = \frac{1}{(2\pi)^{\frac{3}{2}}} e^{i\mathbf{p}\cdot\mathbf{r}} e^{i\Phi(\mathbf{p}, t)} \text{ with } \Phi(\mathbf{p}, t) = \int_0^t \mathrm{d}\tau \left(\mathbf{p} - \frac{\mathbf{A}(t)}{c}\right)^2 \tag{12}
$$

are eigenstates of the momentum operator and form a complete set. This means that in $\bar{\Omega}$ we can expand the KS orbitals as a superposition Volkov waves

$$
\varphi_i(\mathbf{r}, t) = \int \mathrm{d}\mathbf{p}\, b_i(\mathbf{p}, t) \chi_{\mathbf{p}}(\mathbf{r}, t) \tag{13}
$$

and that, from (8) and (7), the number of escaped electrons can be written as

$$
N_{\mathrm{esc}}(t) = 2\sum_{i=1}^{N/2} \int_{\bar{\Omega}} \mathrm{d}\mathbf{r}\, |\varphi(\mathbf{r}, t)|^2 = 2\sum_{i=1}^{N/2} \int \mathrm{d}\mathbf{p}\, |b_i(\mathbf{p}, t)|^2. \tag{14}
$$

This provides us with an expression for $N_{\text{esc}}(t)$ with an explicit dependence on $\mathbf{p}$ that we can exploit to calculate the the resolved probability from (9)

$$\mathscr{P}(\mathbf{p}, t) = \frac{1}{N} \frac{\partial N_{\text{esc}}(t)}{\partial \mathbf{p}} = \frac{2}{N} \sum_{i=1}^{N/2} |b_i(\mathbf{p}, t)|^2 \tag{15}$$

from which the connection with the Volkov expansion coefficients $b_i(\mathbf{p}, t)$ is apparent. This result is consistent with the construction of the Hamiltonian (10) which serves the purpose of exposing the momentum dependence of the scattering states by approximating them with Volkov waves and offers the interpretation of the expansion coefficients as scattering amplitudes.

The second part of the problem is the evaluation $b_i(\mathbf{p}, t)$ from the knowledge of the TDKS orbitals. To this end we express $N_{\text{esc}}(t)$ as a flux integral of the current density $\mathbf{J}(\mathbf{r}, t)$. By applying the continuity equations for the volume $\bar{\Omega}$, we find that

$$N_{\text{esc}}(t) = -\int_0^t d\tau \oint_S d\mathbf{s} \cdot \mathbf{J}(\mathbf{r}, \tau), \tag{16}$$

where $d\mathbf{s}$ is the surface element vector pointing in the direction normal to the surface $S$ enclosing the region $\Omega$ as shown in Fig. 2a. The current density can be evaluated as an expectation value over the TDKS wavefunction as

$$\mathbf{J}(\mathbf{r}, t) = 2 \sum_{i=1}^{N/2} \langle \varphi_i(t)|\hat{\mathbf{j}}|\varphi_i(t)\rangle \text{ with } \hat{\mathbf{j}} = \text{Re}\left\{-i\nabla - \frac{\mathbf{A}(t)}{c}\right\} \tag{17}$$

being the single-particle current operator. If we expand the bra $|\varphi_i(t)\rangle$ in Volkov waves with (13), we have that

$$\mathbf{J}(\mathbf{r}, t) = 2 \sum_{i=1}^{N/2} \int d\mathbf{p}\, b_i(\mathbf{p})\langle \varphi_i(t)|\hat{\mathbf{j}}|\chi_{\mathbf{p}}(t)\rangle, \tag{18}$$

together with the conjugate expression obtained by expanding $\langle \varphi_i(t)|$. By plugging this equation and its conjugate into (16), we obtain a formula that we can directly confront to (14). From the comparison, it results that the Volkov expansion coefficients can be written as the surface flux integral of the projected single-particle current operator $\hat{\mathbf{j}}_{\mathbf{p},i}(\mathbf{r}, t) \equiv \langle \chi_{\mathbf{p}}(t)|\hat{\mathbf{j}}|\varphi_i(\tau)\rangle$ accumulated over time

$$b_i(\mathbf{p}, t) = -\int_0^t d\tau \oint_S d\mathbf{s} \cdot \langle \chi_{\mathbf{p}}(t)|\hat{\mathbf{j}}|\varphi_i(\tau)\rangle. \tag{19}$$

This formulation is particularly convenient since it only involves the evaluation of $\hat{\mathbf{j}}_{\mathbf{p},i}(\mathbf{r}, t)$ on the surface $S$ and, given

$$\hat{\mathbf{j}}_{\mathbf{p},i}(\mathbf{r}, t) = \frac{1}{2} \left\{ i\varphi_i(\mathbf{r}, t)\nabla\chi_{\mathbf{p}}^*(\mathbf{r}, t) - i\chi_{\mathbf{p}}^*(\mathbf{r}, t)\nabla\varphi_i(\mathbf{r}, t) - 2\frac{\mathbf{A}(t)}{c}\chi_{\mathbf{p}}^*(\mathbf{r}, t)\varphi_i(\mathbf{r}, t) \right\},$$

(20)

it is easy to see that it requires the evaluation of the gradient operator only on $S$. Furthermore, as long as the time integral runs for times longer than the complete ionization process, during the time propagation, electrons can seamlessly flow back and forth through the surface without producing any artifact in the final spectrum. This provides with the freedom to choose the radius of $\Omega$ only on basis of the quality of the approximation of (10) which in practice has to be converged case by case.

## 4     Photoemission from Crystal Surfaces

The t-SURFF formalism can be easily extended to include the description of electron photoemission from bulk crystals. Since the derivation largely overlaps with the one presented in the previous section, here we focus on the differences and remind the interested reader to De Giovannini et al. (2017) for the full derivation.

A crystal can be considered as a very large molecule with a highly ordered lattice structure; thus, in principle, one just needs to find a suitable geometry to calculate the flux of the ionization current. The need for a surface to calculate the flux immediately implies that the model must include the material termination from which the electrons can escape. Therefore, in contrast to the usual practice of solid-state physics where the crystal is considered as a periodically repeating lattice in all directions, we need to reduce the periodicity from three to two dimensions and explicitly introduce a surface. Given this arrangement, one can choose $S$ as an infinite plane oriented perpendicular to the nonperiodic dimension like in Fig. 2b and calculate the photoelectron probability with (15) and (19).

Further, to exploit the translational symmetry, one can use the periodic formulation of the TDKS equations (6) and (7) on a bidimensional primitive cell with lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ spanning the surface. At this point we just need to adapt the t-SURFF equations to the new periodic boundary conditions.

These boundary conditions impose slight but crucial modifications to the Volkov waves. First, applying the Bloch theorem to the Volkov Hamiltonian results into a decomposition of the momentum $\mathbf{p}$ in the definition of (12). Indeed, given the Bloch factorization (5), the plane-wave momentum $\mathbf{p}$ can be seen as a sum of a bidimensional crystal momentum $\mathbf{k}$ plus a reciprocal lattice vector $\mathbf{G}$ – a component of the spatial Fourier transform on the unit cell: $\mathbf{p} = \mathbf{k} + \mathbf{G}$. Because of periodicity, the component of $G$ parallel to the surface, $\mathbf{G}_{\parallel}$, can assume only discrete values, while the perpendicular one, $\mathbf{G}_{\perp}$, is continuous. Second, $\chi_{\mathbf{p}}(\mathbf{r}, t)$ must be normalized to the unit cell area such that the Volkov wave definition (12) becomes

$$\chi_{\mathbf{p}}(\mathbf{r}, t) = \sqrt{\frac{2\pi}{a_1 a_2}} e^{i\mathbf{p}\cdot\mathbf{r}} e^{i\Phi(\mathbf{p}, t)} = \sqrt{\frac{2\pi}{a_1 a_2}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} e^{i\Phi(\mathbf{k}+\mathbf{G}, t)},$$

(21)

while the definition of the phase $\Phi(\mathbf{p}, t)$ is unmodified.

With this prescription, we can expand each KS orbital for a given $\mathbf{k}$ with Volkov waves at the same $\mathbf{k}$. The Volkov expansion given by (13) turns into

$$\phi_{i,\mathbf{k}}(\mathbf{r}, t) = \int d\mathbf{G}\, b_i(\mathbf{k} + \mathbf{G}, t)\chi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}, t), \tag{22}$$

where the decomposition of the momentum $\mathbf{p}$ explicitly appears.

The expansion coefficients $b_i(\mathbf{p}, t)$ even in this case act as scattering amplitudes and are central to the definition of $\mathscr{P}(\mathbf{p}, t)$. In fact using the expansion (22), it is possible to derive the same expression for the photoelectron probability given by (15). The difference in the derivation is that, likewise for the definition of the charge density of a periodic system (7), the equations for the number of escaped electrons $N_{\mathrm{esc}}(t)$ (14) and the current density $\mathbf{J}(\mathbf{r}, t)$ (18) have to be integrated over the full BZ. Once the connection between photoelectron probability and expansion coefficients is established, we are left with the problem of calculating the coefficients.

Like in the nonperiodic case, the expansion coefficients can be defined with an expression involving a flux integral of the current density. This results in an equation identical to (19), only for $\phi_{i,\mathbf{k}}(\mathbf{r}, t)$. This formula can be further simplified using the Bloch factorizations of $\chi_{\mathbf{p}}(\mathbf{r}, t)$ and $\phi_{i,\mathbf{k}}(\mathbf{r}, t)$ to obtain an equation containing only the periodic components of the orbitals $u_{i,\mathbf{k}}(t)$ and plane waves $|\mathbf{G}\rangle$ with the periodicity of the lattice

$$b_i(\mathbf{p} = \mathbf{k} + \mathbf{G}, t) = -\int_0^t d\tau \oint_S d\mathbf{s} \cdot \langle \mathbf{G}|\hat{\mathbf{j}}|u_{i,\mathbf{k}}(t)\rangle e^{i\Phi(\mathbf{k}+\mathbf{G}, t)}. \tag{23}$$

This final expression is well suited for the periodic TDKS equations (6) and (7) and any numerical implementation designed to solve them.

## 5   Applications

Nowadays TDDFT can be used to calculate the photoelectron spectrum of a large variety of material complexes and laser configurations. Given the general form in which the field $\mathbf{A}(t)$ enters the TDKS equations, the formalism trivially includes the special case where the system is simultaneously irradiated by a pump and a probe laser pulses with a given time delay, $\tau$, just by employing a field of the form

$$\mathbf{A}(t) = \mathbf{A}_{\mathrm{pump}}(\omega_P, t) + \mathbf{A}_{\mathrm{probe}}(\omega, t + \tau). \tag{24}$$

Likewise in the experiments where, in order to expand the dependence on $\tau$, one has to repeat the measure for every value of the delay, in the simulations, one needs to perform a separate calculation for every different configuration of pulses.
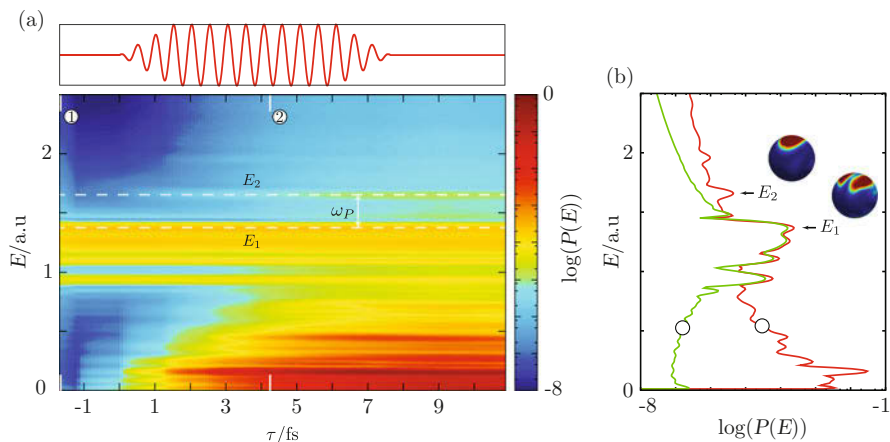
**Fig. 3** Pump-probe photoelectron spectrum of ethylene. The time-resolved photoelectron spectrum (**a**) is calculated by performing a different simulation for each value of $\tau$. The spectra at fixed time delays (**b**) can be further analyzed expanding the angular dependence of electrons ionized with a fixed kinetic energy displayed as spheres in the inset. (Reproduced from De Giovannini et al. 2013)

As an example, consider the case of an ethylene molecule pumped by a laser pulse resonant with the energy of transition between the HOMO – the highest occupied molecular orbital – and the LUMO, the lowest empty molecular orbital. The plot in Fig. 3a is obtained by calculating several energy-resolved photoelectron spectra $\mathscr{P}(E)$ and illustrates how the resonant excitation takes place as function of $\tau$. The appearance of a population on the excited state is clearly observable as an increase of the photoelectron probability for the energy $E_2$ – corresponding to the LUMO energy – which survives for delay times longer than the pump pulse.

Having access to the full $\mathscr{P}(\mathbf{p})$ provides a higher level of detail that can be used to further inspect the excitation process. For instance, for a fixed $\tau$, one can analyze the angular distribution of the photoelectrons emerging with a given kinetic energy: $\mathscr{P}(p = \sqrt{2E}, \theta, \phi)$. The plot in Fig. 3b shows how by applying this analysis to the HOMO and LUMO energies, $E_1$ and $E_2$, it is possible to observe the correlation between the nodal structure of the spatial orbitals and the angular distribution of the photoelectrons.

So far we discussed molecular models where the only role of the nuclei is to provide the source of the external potential, but a full molecular problem involves the entire electron-nuclear wavefunction. The simplest strategy is the so-called Ehrenfest dynamics in which the ions are fully classical and are coupled to the electrons via the density (see plenary ▶ Chap. 6, "TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter"). To illustrate how the ion dynamics can reflect in the photoelectron spectrum in Fig. 4, we present the time-resolved photoelectron spectrum of ethylene initialized into an excited state where one electron is promoted from the HOMO to the LUMO. Given
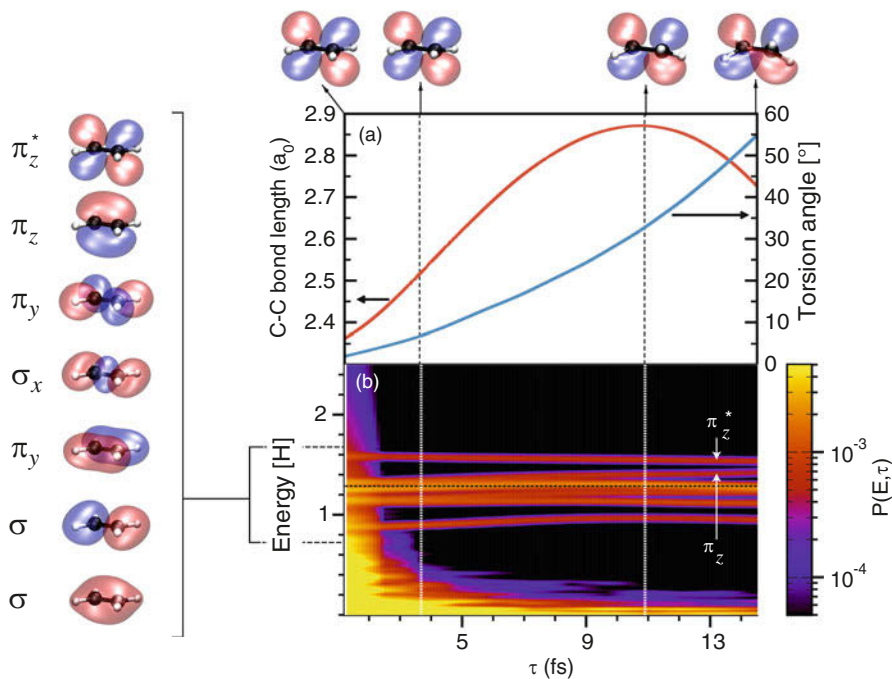
**Fig. 4** Pump-probe photoelectron spectrum of ethylene with moving ions. At time zero the system is initialized by promoting one electron from the HOMO ($\pi_z$ bonding) to the LUMO ($\pi_z^*$ antibonding), and the evolution is monitored by probing the system at different time delays. The time evolution of the molecular geometry (**a**) is reflected in the time-resolved spectrum (**b**). (Reproduced from Crawford-Uranga et al. 2014)

the antibonding nature of the LUMO orbital – of $\pi^*$ symmetry – the excitation initiates an ionic motion on the C-C bond which results in the combination of a vibrational and a torsional mode shown in Fig. 4a. This nuclear motion is reflected in the electronic configurations, and their time evolution can be monitored with time-resolved PES by probing the system at different time delays as in Fig. 4b. In particular, one can observe how the peaks corresponding to the HOMO, $\pi_z$, and the LUMO, $\pi_z^*$, move in energy toward each other as the torsion angle increases. A fact easy to interpret by looking at the shape of the orbitals, depicted in the inset of the figure, and noticing that in the limit of 90° torsion $\pi_z$ and $\pi_z^*$ become degenerate.

Of course the TDDFT methodology described so far is by no means limited to small molecules. For instance, in Wopperer et al. (2017), the photoelectron spectrum of $C_{60}$ was easily simulated for a single strong-field laser pulse. In practical calculations, the largest part of the computational burden is carried by the amount of vacuum needed to account for ionization, i.e., the size of the region $\Omega$. This computational cost scales with the third power of the radius of $\Omega$, while the TDKS equations have a much lighter linear scaling with the total number of electrons $N$.
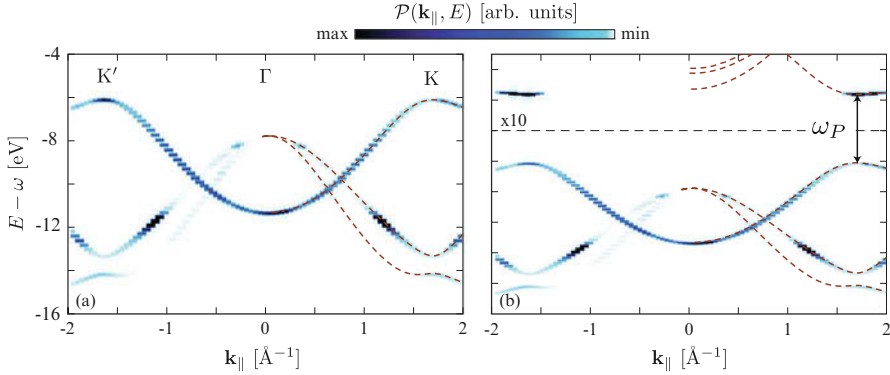
**Fig. 5** Pump-probe ARPES on monolayer hBN. The ARPES spectrum (**a**) of the system in equilibrium is modified by pumping it with a laser pulse resonant with the gap at K. The resulting pump-probe ARPES spectrum (**b**) displays finite electron populations of the electrons in the conduction band across the gap. The band structure is plotted as dashed lines. (Reproduced from De Giovannini et al. 2017)

Therefore the size of the system itself is not critical for a relatively large number of atoms ($\approx$50).

The scaling with the volume of $\Omega$ is less severe for periodic systems. Indeed, since the vacuum only extends in one dimension, the computational cost scales linearly. In this kind of calculations however, the number of points needed to sample the reciprocal space – the k-points – constitutes an additional dimension that one needs to take into account. Thus even though the primitive cell contains only few atoms, the cost of the simulation can be considerable. Nevertheless the simulation of pump-probe photoelectron spectra for periodic systems can be reduced to a manageable task.

As an example in Fig. 5, we present the photoelectron spectrum of monolayer hexagonal boron-nitride (hBN). The ARPES spectrum can be directly obtained from the photoelectron distribution as $\mathscr{P}(\mathbf{p}_\parallel, E)$. This is in perfect agreement with the band structure shown in Fig. 5a. The ARPES spectrum is modified by pumping the system with a pulse resonant with the gap and by probing it after the pump has been switched off, Fig. 5b. Granted the fact that ARPES probes only occupied states, the appearance of a signal in correspondence of the conduction band at K is a clear indication of the population transfer operated by the pump.

For the sake of clarity, in this chapter, we restricted the discussion to a simplified formulation, and it is however important to keep in mind that the general theory can handle more complex systems than simple nonmagnetic insulators. For instance, it can easily describe systems with non-negligible spin-orbit coupling, such as monolayer $WSe_2$, and in highly out-of-equilibrium configurations, like the example in Fig. 6. Similar to the former case, the system is pumped with a laser resonant with the bandgap at K but is probed at different time delays in such a way to explore different overlap configurations between the pump and the probe fields.
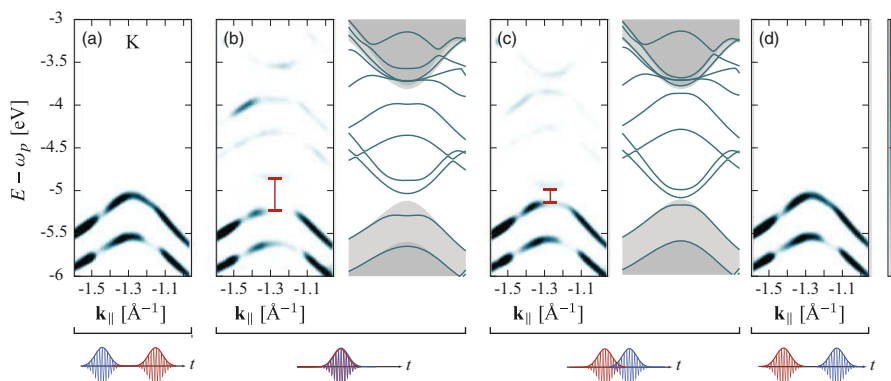
**Fig. 6** Pump-probe ARPES on monolayer WSe$_2$ strongly driven out of equilibrium. The spectra are obtained with different pump-probe configurations: (**a**) the probe comes before the pump, (**b**) the probe is fully contained in the pump, (**c**) the pump and the probe only partially overlap, and (**d**) the probe comes after the pump. The photon-dressed bands are reported for comparison in panels (**b**) and (**c**). Pump (red) and probe (blue) configurations are depicted as cartoons below each panel. (Reproduced from De Giovannini et al. 2016)

When the pump field is strongly coupled with the electrons, the system can undergo a transition into a state where quasiparticles formed by the dressing of photons to electrons emerge as new eigenstates. This is typically accompanied by a proliferation of the energy levels like the one observed in the Autler-Townes effect in atomic physics. The same dressing mechanism can be observed in Fig. 6b, c where the ARPES signal in correspondence of the conduction band becomes deformed and splits in two – this feature is indicated by the red marker bar in the figure. This is a highly out-of-equilibrium feature that can be properly understood in the context of the Floquet theory – for details, see De Giovannini et al. (2016).

## 6    Conclusions

From the standpoint of the experiments, pump-probe photoelectron spectroscopy nowadays appears as a mature field where the key tools are established and have been tested and available in many laboratories around the world. The perspective of the field is the application of this toolset to investigate new dynamical processes in a vast array of system going from atoms to large molecules and bulk crystals.

This roadmap raises important challenges for the theory, some of which can be tackled within TDDFT. Using the versatility of TDDFT, the goal of this chapter was to present an overview of this vast field by covering finite and infinite systems on the same ground.

Clearly, the accuracy in the description of the photoemission process presented here is affected by the same restrictions involved with the approximation of the exchange and correlation functionals in TDDFT. Future developments aimed at

improving these functionals to overcome present limitations, such as the description of static and dynamic correlations or long-range interactions, will positively affect photoelectron observables. Much-needed developments in this direction will make possible the description of exciton dynamics in solids or double-excitation phenomena such as Fano resonances which currently lay beyond the reach of the application of the theory.

We conclude mentioning that the code to perform real-space and real-time TDDFT simulations of pump-probe photoelectron spectra with the techniques presented in this chapter is freely available to the public (Andrade et al. 2015).

# References

Andrade X, Strubbe D, De Giovannini U, Larsen AH, Oliveira MJT, Alberdi-Rodriguez J, Varas A, Theophilou I, Helbig N, Verstraete MJ, Stella L, Nogueira F, Aspuru-Guzik A, Castro A, Marques MAL, Rubio A (2015) Real-space grids and the octopus code as tools for the development of new simulation approaches for electronic systems. Phys Chem Chem Phys 17(47):31371–31396

Bachau H, Cormier E, Decleva P, Hansen JE, Martín F (2001) Applications of B-splines in atomic and molecular physics. Rep Prog Phys 64(12):1815

Blaga CI, Catoire F, Colosimo P, Paulus GG, Muller HG, Agostini P, Dimauro LF (2009) Strong-field photoionization revisited. Nat Phys 5(5):335

Castro A, Marques MAL, Rubio A (2004) Propagators for the time-dependent Kohn–Sham equations. J Chem Phys 121(8):3425

Catoire F, Bachau H (2012) Extraction of the absolute value of the photoelectron spectrum probability density by means of the resolvent technique. Phys Rev A 85(2):138

Crawford-Uranga A, De Giovannini U, Mowbray DJ, Kurth S, Rubio A (2014) Modelling the effect of nuclear motion on the attosecond time-resolved photoelectron spectra of ethylene. J Phys B-At Mol Opt 47(12):124018

Damascelli A, Hussain Z, Shen ZX (2003) Angle-resolved photoemission studies of the cuprate superconductors. Rev Mod Phys 75(2):473

De Giovannini U, Brunetto G, Brunetto G, Castro A, Walkenhorst J, Walkenhorst J, Rubio A (2013) Simulating pump-probe photoelectron and absorption spectroscopy on the attosecond timescale with time-dependent density functional theory. Chem Phys Chem 14(7):1363–1376

De Giovannini U, Larsen AH, Rubio A (2015) Modeling electron dynamics coupled to continuum states in finite volumes with absorbing boundaries. Eur Phys J B 88(3):1

De Giovannini U, Hübener H, Rubio A (2016) Monitoring electron-photon dressing in WSe2. Nano Lett 16:7993

De Giovannini U, Hübener H, Rubio A (2017) A first-principles time-dependent density functional theory framework for spin and time-resolved angular-resolved photoelectron spectroscopy in periodic systems. J Chem Theory Comput 13:265

Hsieh D, Basov DN, Averitt RD (2017) Towards properties on demand in quantum materials. Nat Mater 16(11):1077

Kadanoff LP, Baym G (1962) Quantum statistical mechanics. W. A. Benjamin, New York

Krausz F, Stockman MI (2014) Attosecond metrology: from electron capture to future signal processing. Nat Photon 8(3):205

Legrand C, Suraud E, Reinhard PG (2002) Comparison of self-interaction-corrections for metal clusters. J Phys B-At Mol Opt 35(4):1115

Lépine F, Ivanov MY, Ivanov MY, Vrakking MJJ, Vrakking MJJ (2014) Attosecond molecular dynamics: fact or fiction? Nat Photon 8(3):195

Pendry JB (1990) Low-energy electron diffraction. In: Bortolani V, March NH, Tosi MP (eds) Interaction of atoms and molecules with solid surfaces. Phys of Solids Liq. Springer, Boston

Perfetto E, Sangalli D, Marini A, Stefanucci G (2016a) First-principles approach to excitons in time-resolved and angle-resolved photoemission spectra. Phys Rev B 94(24):245303

Perfetto E, Uimonen AM, van Leeuwen R, Stefanucci G (2016b) Time-resolved photoabsorption in finite systems: a first-principles NEGF approach. J Phys Conf Ser 696(1):012004

Puschnig P, Berkebile S, Fleming AJ, Koller G, Emtsev K, Seyller T, Riley JD, Ambrosch-Draxl C, Netzer FP, Ramsey MG (2009) Reconstruction of molecular orbital densities from photoemission data. Science 326(5953):702

Stolow A, Bragg AE, Neumark DM (2004) Femtosecond time-resolved photoelectron spectroscopy. Chem Rev 104(4):1719

Tao L, Scrinzi A (2012) Photo-electron momentum spectra from minimal volumes: the time-dependent surface flux method. New J Phys 14(1):013021

Wopperer P, De Giovannini U, Rubio A (2017) Efficient and accurate modeling of electron photoemission in nanostructures with TDDFT. Eur Phys J B 90(3):1307

Wu G, Hockett P, Stolow A (2011) Time-resolved photoelectron spectroscopy : from wavepackets to observables. Phys Chem Chem Phys 13(41):18447

# Modeling Excited States of Confined Systems

# 15

## Linda Hung and Serdar Öğüt

## Contents

### Abstract

The many-body perturbation theory methods based on the $GW$ approximation and the Bethe-Salpeter equation (BSE) provide a first-principles route to modeling one- and two-particle excitations in a variety of bulk and molecular systems. This chapter reviews the current status of $GW$-BSE methods in the context of confined systems. We describe methods for basis set convergence, which allow sufficient numerical precision for accurate benchmarking of $GW$ and BSE theory and study various theoretical approximations within $GW$. The differences in

L. Hung
Toyota Research Institute, Los Altos, CA, USA
e-mail: linda.hung@tri.global

S. Öğüt (✉)
Department of Physics, University of Illinois at Chicago, Chicago, IL, USA
e-mail: ogut@uic.edu

various flavors of $GW$ and $GW$-BSE, including perturbative, self-consistent, and vertex-corrected implementations, are compared in the context of benchmark sets of $sp$-bonded aromatic molecules and Group IB and IIB transition metal atoms and ions with filled $d$ shells.

# 1    Introduction

From a historical perspective, excited-state descriptions of materials have lagged behind descriptions of the ground state. This is especially true for density functional theory (DFT) methods where the initial formalism was developed to describe only ground-state properties. However, studies during the last three decades on time-dependent density functional theory (TDDFT) and Green's function methods, such as the $GW$ approximation and the Bethe-Salpeter equation (BSE), have provided new methodologies to examine excited-state properties within similar frameworks as the ground state (Hedin 1965; Hybertsen and Louie 1986; Strinati 1988; Rohlfing and Louie 2000; Onida et al. 2002; Casida 2009). Developed in the mid-1960s by Hedin, the $GW$ approximation allows one to solve for the interacting one-particle Green's function, whose poles are associated with vertical ionization potentials and electron affinities as measured in photoemission and inverse photoemission experiments. The BSE, on the other hand, allows one to construct an interacting two-particle correlation function from one-particle Green's functions; the poles of the two-particle Green's function are associated with neutral excitations such as those observed in optical absorption experiments. Developments in TDDFT, also used to simulate neutral excitations, are covered in greater detail in another chapter of this *Handbook* ▶ Chap. 6, "TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter".

Since the overview of $GW$ and BSE methods in the first edition of the *Handbook of Materials Modeling* (Louie and Rubio 2005), many-body perturbation theory methods based on the $GW$-BSE formalism have become standard tools to simulate the excited state of materials. However, while certain theoretical methods underlying the $GW$ approximation and the BSE are well established, the accuracy of such methods in predicting properties of real systems is still being assessed and improved. Recent benchmarks for various bulk materials, nanostructures, and molecules are now available to quantify the trade-off between cost and accuracy for approximations within $GW$ and the BSE. In this chapter, we discuss how theoretical and numerical approximations affect the accuracy of excited-state energies for confined systems, e.g., nanoclusters and molecules.

The physical properties of confined systems can deviate from those of extended systems; specifically, stronger self-interaction effects and less delocalized wave functions would be expected (especially in atoms and small molecules). Therefore, any conclusions from benchmarks of TDDFT, $GW$, and the BSE on the solid state might not be completely applicable to confined systems. Nevertheless, in the past 10 years, multiple $GW$ and $GW$-BSE benchmarks specifically focused on molecules have been reported to quantify the accuracy of different levels of

theory (Ma et al. 2010; Rostgaard et al. 2010; Ke 2011; Faber et al. 2011; Qian et al. 2011; Blase and Attaccalite 2011; Ren et al. 2012; Baumeier et al. 2012; Marom et al. 2012; Bruneval and Marques 2013; Pham et al. 2013; Caruso et al. 2013; Atalla et al. 2013; Koval et al. 2014; Boulanger et al. 2014; Körbel et al. 2014; Hirose et al. 2015; Wang 2015; Krause et al. 2015; Bruneval et al. 2015; Jacquemin et al. 2015; van Setten et al. 2015; Hung et al. 2016, 2017; Knight et al. 2016; Blase et al. 2011, 2016). Several of these benchmarks also compare software packages to validate the different implementations. In general, $GW$-BSE methods have proven to be rather accurate for modeling charged and neutral excitations, particularly for excitations between orbitals with $s$ and $p$ character. For systems containing transition metal elements, however, computations using the $GW$-BSE formalism face greater challenges, since the enhanced electron correlations inherent in such systems necessitate the use of more sophisticated approaches beyond simple perturbative implementations. Understanding the level of theory needed to achieve a reasonable balance between accuracy and computational demand is therefore a crucial step in enhancing the predictive power of the $GW$-BSE formalism and extending its applicability to a wider range of materials.

In Sect. 2, we begin by reviewing the theory and implementation of the $GW$-BSE framework, including descriptions of prevalent approximations. In Sect. 3, we discuss methods that improve the numerical convergence of energies, which enables the theoretical accuracy of $GW$-BSE to be accurately compared. In Sect. 4, we apply various flavors of $GW$-BSE methods, including perturbative (one-shot) $G_0W_0$, $GW$ calculations performed with a simple vertex derived from the local density approximation (LDA) of DFT, and self-consistent $GW$ (in the form of quasiparticle and eigenvalue self-consistency), as well as the ensuing BSE calculations. Our examples are centered on a set of six aromatic molecules comprising benzene ($C_6H_6$), thiophene ($C_4H_4S$), 1,2,5-thiadiazole ($C_2H_2N_2S$), naphthalene ($C_{10}H_8$), benzothiazole ($C_7H_5NS$), and tetrathiafulvalene (TTF, $C_6H_4S_4$) and Group IB and IIB atoms in three valence electron configurations: $d^{10}$ ($Cu^+$, $Ag^+$, $Zn^{2+}$, and $Cd^{2+}$), $d^{10}s^1$ ($Cu^0$, $Ag^0$, $Zn^+$, and $Cd^+$), and $d^{10}s^2$ ($Cu^-$, $Ag^-$, $Zn^0$, and $Cd^0$). We conclude with a brief summary in Sect. 5.

## 2    Many-Body Perturbation Theory and Implementation

In this section, we present the theory behind the $GW$-BSE framework and an implementation that uses a transition-space representation of excited states. The transition-space (spectral) basis is efficient for confined systems due to their discrete energy levels. It also allows the exact integration of $GW$ self-energies, unlike the contour deformation (Godby et al. 1988) and plasmon pole (Hybertsen and Louie 1986) techniques commonly used with a plane-wave basis. The framework described throughout this section is presented in more detail in Tiago and Chelikowsky (2006).

## 2.1    One-Particle Excitations: The $GW$ Approximation

Electron or hole injection into a system (e.g., processes occurring during ionization or anion formation) is considered single-particle excitations. The energies of these processes are associated with the poles of the one-particle Green's function $G$, which can be expressed as the Dyson equation:

$$G(1, 2) = G_0(1, 2) + \int d(34) G_0(1, 3) \Delta \Sigma(3, 4) G(4, 2). \tag{1}$$

Here, $(1) \equiv (\mathbf{r}_1, \sigma_1, t_1)$ is a many-body notation for the spatial, spin, and time coordinates, and $G_0$ is a mean-field Green's function. While $G_0$ traditionally denotes the noninteracting Green's function (from a Hartree calculation), here we use it to indicate any mean-field calculation, expressed in general as DFT with some exchange-correlation potential $V_{xc}$. Equation 1 also includes the term $\Delta \Sigma$, which is the difference between the self-energy $\Sigma$ and the mean-field exchange-correlation potential.

In Hedin's equations (Hedin 1965), the interacting one-particle Green's function can be determined self-consistently with four other equations that define the polarizability $\chi$, screened Coulomb interaction $W$, self-energy $\Sigma$, and vertex function $\Gamma$:

$$\chi(1, 2) = -i \int d(34) G(1, 3) G(4, 1^+) \Gamma(3, 4; 2), \tag{2}$$

$$W(1, 2) = V_H(1, 2) + \int d(34) V_H(1, 3) \chi(3, 4) W(4, 2), \tag{3}$$

$$\Sigma(1, 2) = i \int d(34) G(1, 3) W(4, 1^+) \Gamma(3, 2; 4), \tag{4}$$

$$\Gamma(1, 2; 3) = \delta(1, 2)\delta(1, 3) + \int d(4567) \frac{\delta \Sigma(1, 2)}{\delta G(4, 5)} G(4, 6) G(7, 5) \Gamma(6, 7; 3), \tag{5}$$

where $1^+$ denotes that $t_1 \to t_1 + \eta$ for some positive infinitesimal $\eta$, $V_H$ is the bare Coulomb potential, and $\delta(1,2)$ is the Dirac delta function.

Hedin's equations are too computationally demanding to model realistic systems. Accordingly, an approximation is taken to reduce the three-point vertex function to

$$\Gamma(1, 2; 3) = \delta(1, 2)\delta(1, 3), \tag{6}$$

which removes the need to evaluate a four-point integral. This vertex approximation gives rise to the naming of the $GW$ approximation, since the expression to compute $\Sigma$ becomes an integral over only $G$ and $W$.

For the implementation of a $GW$ self-energy calculation using a quasiparticle, transition-space basis, the self-energy $\Sigma$ is partitioned into two contributions: a bare exchange part $\Sigma^x$ and a correlation part $\Sigma^c$. The bare exchange self-energy matrix element between quasiparticles $j$ and $j'$ can be written as a sum over occupied states:

$$\langle j|\Sigma^x|j'\rangle = -\sum_n^{\text{occ.}} K^x_{njnj'}, \tag{7}$$

where the exchange kernel is

$$K^x_{vcv'c'} = \int d\mathbf{r} \int d\mathbf{r}' \varphi_v(\mathbf{r})\varphi_c(\mathbf{r}) V_H(\mathbf{r},\mathbf{r}')\varphi_{v'}(\mathbf{r}')\varphi_{c'}(\mathbf{r}'), \tag{8}$$

and $\varphi$ are real-valued quasiparticle wave functions. Since only states $j$, $j'$, and occupied states contribute to this finite summation, the exact value of $\Sigma^x$ (given the basis) can always be computed.

In contrast, the energy-dependent $\Sigma^c$ is expressed as a double infinite sum over quasiparticles $n$ and transitions $s$:

$$\langle j|\Sigma^c(E)|j'\rangle = 2\sum_n^{\infty}\sum_s^{\infty} \frac{V^s_{nj} V^s_{nj'}}{E - \varepsilon_n - \omega_s \eta_n}, \tag{9}$$

where $\eta_n$ is $-1$ for occupied state $n$ (quasihole) and $+1$ for empty state $n$ (quasielectron), and

$$V^s_{nj} = \sum_v^{\text{occ.}}\sum_c^{\text{empty}} K^x_{njvc} \left(\frac{\varepsilon_c - \varepsilon_v}{\omega_s}\right)^{1/2} Z^s_{vc}, \tag{10}$$

where $\varepsilon_n$ are quasiparticle eigenvalues (poles of $G$), $v$ is the index over occupied states, and $c$ is the index over empty states. The transition energies $\omega_s$ (poles of $W$) and eigenvectors with components $Z^s_{vc}$ are from the solution of Casida's equations (Casida 2009).

A cutoff in the infinite sums (over the number of empty states and transitions) therefore must be chosen when computing $\Sigma^c$. As is discussed in Sect. 3, great care must be taken to ensure numerical convergence of the correlation self-energy.

Once $\Sigma^x$ and the energy-dependent $\Sigma^c$ are summed to obtain the total self-energy, the quasiparticle energy $E^{\text{QP}}$ is determined from the Dyson equation for $G$ (Eq. 1), expressed in the quasiparticle basis as

$$\left[-\frac{1}{2}\nabla^2 + V_H(\mathbf{r}) + V_0(\mathbf{r}) + \Delta\Sigma(E^{\text{QP}}_j)\right]\varphi^{\text{QP}}_j = E^{\text{QP}}_j \varphi^{\text{QP}}_j, \tag{11}$$

where $V_0$ is a local external potential with contributions from the ionic potential and the mean-field exchange-correlation potential. In the case of self-consistency, we solve for the $E^{\text{QP}}$ that satisfies Eq. 11 at each iteration. Depending on the level of

self-consistency, the quasiparticle wave functions $\varphi^{\text{QP}}$ may also be updated, but in a separate step. In that case, the full self-energy matrix – not just the diagonal terms – must be computed at each iteration.

### 2.1.1 Vertex Approximations

Multiple works have addressed the issue of improving the description of the vertex beyond that of conventional $GW$ (Eq. 6) while still maintaining a computationally efficient form. These alternate vertex functions may include higher-order terms (Ummels et al. 1998; Maebashi and Takada 2011; Grüneis et al. 2014; Stefanucci et al. 2014) or capture the response of density functionals (Del Sole et al. 1994; Marini and Rubio 2004; Bruneval et al. 2005). In this chapter, we include benchmarks using a vertex derived from a local density functional approach, in which the polarizability $\chi$ is expressed within TDDFT and a consistent level of approximation is achieved by replacing the vertex function with

$$\Gamma_{\text{LDA}}(1, 2; 3) = \delta(1, 2)\delta(1, 3) - i\delta(1, 2) f_{\text{xc}}(1) \times$$
$$\int d(45) G(1, 4) G(5, 1^+) \Gamma_{\text{LDA}}(4, 5; 3), \tag{12}$$

where $f_{xc} = \delta V_{xc}/\delta\rho$ and $V_{xc}$ is the LDA exchange-correlation potential. Although a three-point vertex is needed to accurately describe certain physical properties (Romaniello et al. 2009; Grüneis et al. 2014), this two-point form of the vertex allows computations that cost only a prefactor more than conventional $GW$.

### 2.1.2 Levels of Self-Consistency

Because of the computational cost, $GW$ predictions often are presented as a perturbative (one-shot) result, which is computed by adding only one cycle through Hedin's equations onto the initial mean-field starting point instead of a fully self-consistent procedure. This amounts to solving Eq. 11 for quasiparticle energies only, with quasiparticle wave functions set to the mean-field (e.g., DFT) wave functions. With this so-called $G_0W_0$ approximation, there is a significant dependence of the predicted quasiparticle energies on the starting mean-field description.

Even when a self-consistent result is reported, the level of self-consistency must be specified. Some self-consistent calculations are only eigenvalue self-consistent (ev$GW$), where the "diagonal approximation" is applied so that quasiparticle wave functions remain fixed at the mean-field wave functions and only eigenvalues are updated. A higher level of self-consistency is achieved with quasiparticle self-consistent $GW$ (QS$GW$), where the wave functions are optimized as well. In QS$GW$ (Faleev et al. 2004), the $GW$ self-energy is used to construct a better approximation for the mean-field effective potential, which is in turn used to build a new set of mean-field Green's functions. Using the new Green's functions, the cycle (including updates to $W$) is repeated until self-consistency is reached between the mean-field and new $G_0W_0$ Green's functions. Note that even though QS$GW$ allows

updates to the wave functions, it is different from full self-consistency. Both ev$GW$ and QS$GW$ results are presented in Sect. 4.

## 2.2   Two-Particle Excitations: $GW$-BSE and TDDFT

The promotion of an electron into an excited state is considered a two-particle excitation, due to the coexistence of both an electron and a hole. The BSE expresses the two-particle correlation function $L$ as (Strinati 1988)

$$L(1, 2; 1', 2') = G(1, 2')G(2, 1') + \int d(3456)G(1, 3)G(4, 1')$$

$$K(3, 5; 4, 6)L(6, 2; 5, 2'), \tag{13}$$

where $G$ is, as before, the one-particle Green's function, and the electron-hole interaction kernel is expressed as

$$K(3, 5; 4, 6) = -i\delta(3, 4)\delta(5, 6)V_H(3, 6) + \frac{\delta\Sigma(3, 4)}{\delta G(6, 5)}. \tag{14}$$

Assuming that $G$ is represented using quasiparticles and that electron-hole excited states are long-lived, the BSE can be written as a generalized eigenvalue problem with block matrix form (Rohlfing and Louie 2000):

$$\begin{pmatrix} A & B \\ -B & -A \end{pmatrix} \begin{pmatrix} X_l \\ Y_l \end{pmatrix} = \Omega_l \begin{pmatrix} X_l \\ Y_l \end{pmatrix}, \tag{15}$$

where $\Omega_l$ is the energy of electron-hole excitation $l$.

From the definition of the BSE kernel (Eq. 14), it is clear that $G$ determines the quality of ensuing BSE predictions. When $G$ is obtained from the $GW$ approximation (as in the $GW$-BSE framework) and neglecting dynamical effects, the BSE kernel can be split into an exchange part $K^x$ (Eq. 8) and a direct part:

$$K^d_{vcv'c'} = K^x_{vv'cc'} + 4\sum_s \frac{V^s_{vv'} V^s_{cc'}}{\omega_s}, \tag{16}$$

and the block submatrices for $GW$-BSE corresponding to the spin-conserving excitations (and ignoring spin-orbit interactions) are

$$A_{\uparrow\uparrow,\uparrow\uparrow} = A_{\downarrow\downarrow,\downarrow\downarrow} = D + K^x + K^d,$$

$$B_{\uparrow\uparrow,\uparrow\uparrow} = B_{\downarrow\downarrow,\downarrow\downarrow} = K^x + K^d,$$

$$A_{\uparrow\uparrow,\downarrow\downarrow} = B_{\uparrow\uparrow,\downarrow\downarrow} = A_{\downarrow\downarrow,\uparrow\uparrow} = B_{\downarrow\downarrow,\uparrow\uparrow} = K^x, \tag{17}$$

where

$$D_{vcv'c'} = (E_c - E_v)\delta_{cc'}\delta_{vv'}, \tag{18}$$

with $c$, $c'$ being indices for empty states, $v$, $v'$ being indices for occupied states, and $E_c$ and $E_v$ being their quasiparticle energies. If the ground state is not spin polarized, its neutral singlet (and triplet) excitations can be computed with a basis set two times reduced. Singlet excitations correspond to a BSE Hamiltonian with

$$
\begin{aligned}
A &= D + 2K^x + K^d, \\
B &= 2K^x + K^d.
\end{aligned} \tag{19}
$$

In addition, when a vertex function is used in $GW$ calculations, a corresponding vertex contribution must also be added to the BSE kernel to maintain a consistent level of theory. The LDA vertex term

$$K^f_{vcv'c'} = 2\sum_s \frac{V^s_{vv'}F^s_{cc'} + F^s_{vv'}V^s_{cc'}}{\omega_s}, \tag{20}$$

is added to the BSE Hamiltonian wherever $K^d$ contributes for $GW\Gamma_{\mathrm{LDA}}$-BSE calculations. Here, $F^s_{nj}$ has the same form as $V^s_{nj}$ (Eq. 10) with $K^x_{njvc}$ replaced by $K^{\mathrm{LDA}}_{njvc}$, where LDA exchange-correlation kernel $K^{\mathrm{LDA}}$ is defined as

$$K^{\mathrm{LDA}}_{vcv'c'} = \int d\mathbf{r}\varphi_v(\mathbf{r})\varphi_c(\mathbf{r})f_{xc}(\mathbf{r})\varphi_{v'}(\mathbf{r})\varphi_{c'}(\mathbf{r}). \tag{21}$$

The BSE can also use a lower level of approximation for $G$, constructing it directly from wave functions and energies ($\varphi$ and $\varepsilon$) corresponding to the Kohn-Sham DFT electronic structure, with the self-energy in the BSE kernel (Eq. 14) approximated by $\Sigma(1,2) \approx V_{\mathrm{xc}}(1)\delta(1,2)$. This form of the BSE is in fact linear-response TDDFT in Casida's formalism. The eigenvalue equation can be written as

$$R^{1/2}\left[R + 4(K^x + K^{\mathrm{LDA}})\right]R^{1/2}Z_s = \omega_s^2 Z_s. \tag{22}$$

Here, the matrix $R$ is diagonal with entries $R_{vcv'c'} = \delta_{vv'}\delta_{cc'}(\varepsilon_c - \varepsilon_v)$. Compared to $GW$-BSE, TDDFT offers a computationally more efficient method for calculating absorption spectra of confined systems ▶ Chap. 6, "TDDFT and Quantum-Classical Dynamics: A Universal Tool Describing the Dynamics of Matter".

As a final note, the BSE frameworks outlined above represent $G$ in the diagonal approximation. However, past work has shown that off-diagonal terms can alter $G_0W_0$ and $G_0W_0$-BSE energies by more than 1 eV, especially for unoccupied states in finite systems or when occupied states are poorly described within DFT (Rohlfing and Louie 2000; Kaplan et al. 2015). In this chapter, we include the off-diagonal terms with the expression

$$D_{vcv'c'} = \delta_{vv'} \sum_{\bar{c}} \varepsilon_{\bar{c}} \langle c|\bar{c}\rangle \langle \bar{c}|c'\rangle$$

$$- \delta_{cc'} \sum_{\bar{v}} \varepsilon_{\bar{v}} \langle v|\bar{v}\rangle \langle \bar{v}|v'\rangle, \tag{23}$$

where $c$ and $v$ are indices for the original Kohn-Sham DFT electronic structure and the overbar indicates energies and wave functions of the diagonalized quasiparticle basis. The equation above is used in place of Eq. 18 when solving the BSE. Note that the expression above corresponds to the case where the occupied-state subspace and the unoccupied-state subspace are diagonalized separately and the electron density remains unchanged (Tiago and Chelikowsky 2006; Hung et al. 2017).

## 3 Numerical Convergence

The typical target for theoretical accuracy in $GW$ and BSE is ∼0.1 eV. Therefore, the necessary numerical accuracy must also be near ∼0.1 eV. This level of accuracy in the computation of charged and neutral excitation energies of confined systems requires one to pay particular attention to convergence issues within the $GW$-BSE framework. We focus here on the evaluation of the correlation self-energy (Eq. 9), where summations over empty states converge slowly and become the bottleneck of the computation. More generally stated, this is where convergence relative to basis set size is most difficult. The challenge in obtaining the convergence of excited-state energies with basis set size has motivated the development and testing of techniques to accelerate convergence or approximate the complete basis set limit. These include extrapolation (Sharifzadeh et al. 2012; Klimeš et al. 2014; Hung et al. 2014), the extrapolar method (Anglade and Gonze 2008; Bruneval and Gonze 2008; Berger et al. 2010), Lanczos-chain techniques (Rocca et al. 2008; Umari et al. 2010), construction of more efficient basis sets (Samsonidze et al. 2011; Jiang and Blaha 2016), sampling fewer high-energy states (Gao et al. 2016), and approximations for the missing basis set contributions (Tiago and Chelikowsky 2006; Kang and Hybertsen 2010; Deslippe et al. 2013).

The methods that approximate missing basis set contributions are based on adding a "static remainder" correction to the $GW$ correlation self-energy. This correction is derived from the static limit of the $GW$ approximation, which is known as the Coulomb-hole-screened-exchange (COHSEX) approximation, and is added to summation over quasiparticles (Eq. 9) truncated at some quasiparticle index $N$. In work by Tiago and Chelikowsky (2006), the COHSEX correction is given by the difference between the total COHSEX $\Sigma^c$ calculated using a completeness relation and the COHSEX $\Sigma^c$ in sum-over-states form, truncated at the same index $N$. More recent work shows that a correction of half that difference is a better approximation (Kang and Hybertsen 2010; Deslippe et al. 2013).

While the static remainder already improves convergence for most confined systems, it is important to note that it only enhances the convergence of the

summation over quasiparticle states. In addition, numerical convergence must also be sought with respect to the infinite summation over transitions $s$ in the expression for $\Sigma^c$. We focus here on methods of extrapolation to achieve that convergence.

To approach the complete basis set limit with respect to transitions $s$ via extrapolation, one can calculate the $GW$ energies with increasing number of quasiparticle states $N$ and increasing number of transitions $s$ between those states. The converged $GW$ energy is then obtained by fitting $E_\infty$ and coefficients $c_1$ and $c_2$ to

$$E(N) = E_\infty + \frac{c_1}{N + c_2}, \tag{24}$$

where the $1/N$ dependence has been suggested (Gulans 2014) as arising from the electron-electron cusp condition. With atom-centered Gaussian-type basis sets that are more common in quantum chemistry methods, similar extrapolation and fitting schemes can be used. Since all quasiparticles and transitions that exist within a given basis are included in the computation, extrapolation is performed across basis sets of increasing size. With Dunning basis sets, the $GW$ energies at the complete basis set limit are obtained by fitting the extrapolated energy $E_\infty$ and coefficient $c_1$ to

$$E(X) = E_\infty + c_1 X^{-\alpha} \tag{25}$$

where $X = 3, 4$, or $5$ ranging from aug-cc-PVTZ, aug-cc-PVQZ, or aug-cc-PV5Z, respectively. An $X^{-3}$ dependence – equivalent to the $N^{-1}$ scaling where $N$ is the number of basis functions – is commonly used for extrapolating the correlation energy with Dunning basis sets (Bruneval 2012; Kaplan et al. 2016). However, a smaller exponent ($\alpha \approx 2$) has been empirically shown to provide a better fit for quasiparticle energies computed on finite basis sets (Truhlar 1998; Bruneval et al. 2016).

In this chapter, the majority of computations use the RGWBS software suite (Tiago and Chelikowsky 2006), with numerical convergence achieved with a combination of the half COHSEX remainder and extrapolation. Results from BerkeleyGW (Deslippe et al. 2012) are converged with a combination of sampling high-energy states and extrapolation. Results from MolGW (Bruneval et al. 2016) use extrapolation.

## 4 Simulating Atoms and Molecules

### 4.1 Numerical Validation

Software packages make a variety of choices in their numerical implementation that may affect the final computed $GW$ results. We compare the RGWBS software suite with BerkeleyGW and MolGW and find that the numerical error is fairly well controlled. RGWBS, which represents wave functions on a uniform real-space grid,

**Fig. 1** Error of $G_0W_0$ IPs relative to experiment. Computations use two software packages: black circles are RGWBS (real-space framework), and orange squares are BerkeleyGW (plane-wave framework)
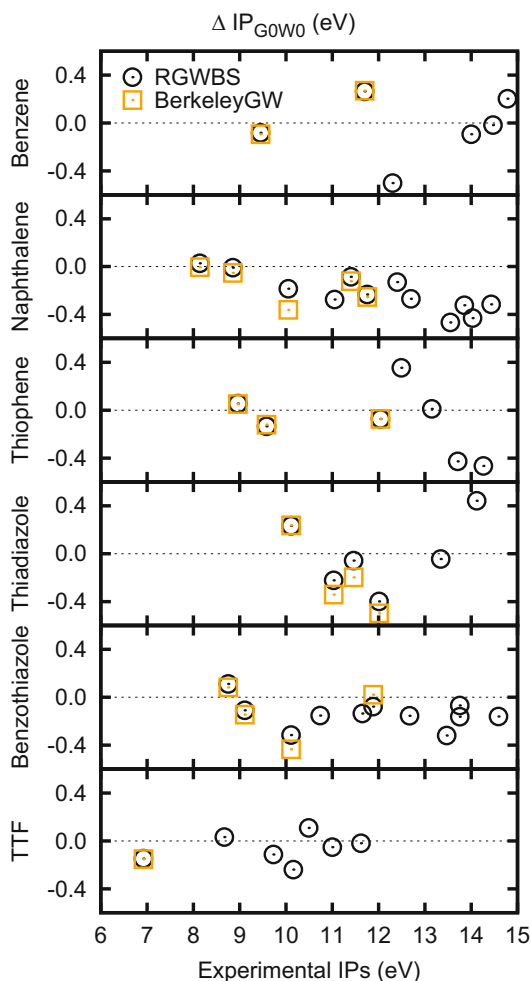


and BerkeleyGW, which has a plane-wave basis, are compared in Fig. 1 across the aromatic molecule test set (Hung et al. 2016). The frontier orbitals differ by no more than 80 meV, and only five IPs differ by more than 100 meV despite the different basis sets and other numerical approximations. A similar comparison has also been made comparing MolGW, which represents wave functions with Gaussian bases, with RGWBS for transition metal atoms and ions, with differences at the complete basis set limit near ~0.2 eV (Hung et al. 2017).

Figure 1 also shows that, often, $G_0W_0$ with a LDA or generalized gradient approximation (GGA) starting point is sufficient for good agreement with experiment in $sp$-bonded systems (see Sect. 4.3 below). As has also been noted in earlier work, we observe that the error in $G_0W_0$ energies is insensitive to the type of orbital, correcting from DFT where the energies of orbitals with $\sigma$ character are significantly

less accurate than those of $\pi$ orbitals (Hybertsen and Louie 1986; Qian et al. 2011; Marom et al. 2012).

## 4.2    Starting-Point Dependence

Among the various approximations, the effect of the mean-field starting point on the quasiparticle energies of molecules has perhaps been the most studied. In multiple benchmarks, hybrid functional mean-field starting points have been consistently observed to produce better agreement between $G_0W_0$ and experiment as seen in benchmarks referenced in Sect. 1. Here we only briefly touch on one set of examples illustrating the starting-point dependence.

The effect of the amount of exact exchange in the mean-field starting point on $GW$ predictions is shown in Fig. 2 for the Zn atom and ions. Computations in MolGW are performed where none, half, or all of the exchange density functional is replaced with a corresponding amount of exact exchange (Hung et al. 2017). At the DFT level, the lowest unoccupied molecular orbital (LUMO) becomes less bound with increasing amounts of exact exchange, while the highest occupied molecular orbital (HOMO) is more bound. The $G_0W_0$ results decrease or even reverse the trend observed at the DFT level, and the changes in trends are most evident at the ev$GW$
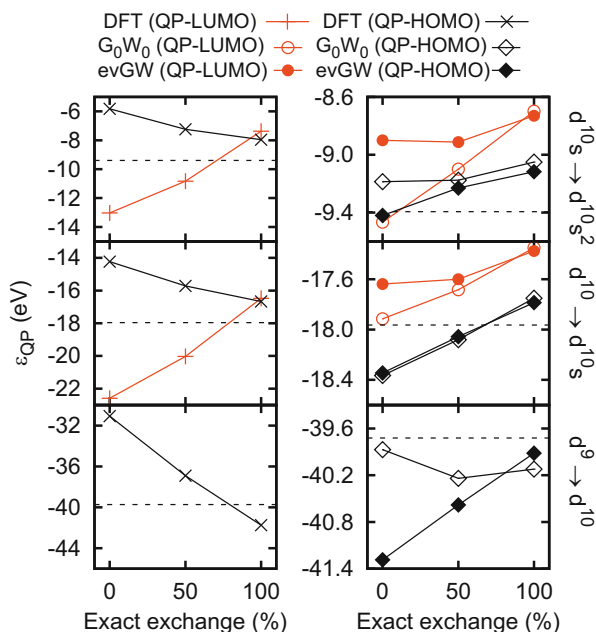


**Fig. 2**  DFT (left), $G_0W_0$, and ev$GW$ (right) quasiparticle energies of Zn, $Zn^+$, and $Zn^{2+}$ with mean-field starting points of 0% exact exchange (GGA), 50% exact exchange, and 100% exact exchange (HF). Dashed lines indicate the experimentally measured IPs

level of theory. This example illustrates that the starting-point dependence persists even if eigenvalues are iterated to self-consistency. The starting-point contribution may only be removed with optimization of the quasiparticle wave functions as well.

## 4.3    Combining $\Gamma$ and Self-Consistency

The LDA-derived vertex approximation has, in the past, been generally associated with a near-constant shift of all energies compared to conventional $GW$ values. We see the same trend for our test set of aromatic molecules (Fig. 3). Compared to $G_0W_0$ ionization potentials (IPs), $G_0W_0\Gamma_{LDA}$ predictions are consistently shifted downward by $\sim 0.7$ eV. The electron affinities (EAs) are similarly shifted downward



**Fig. 3** Shift of quasiparticle energies from predictions at $G_0W_0$, for $GW$ variants including self-consistency and vertex corrections, for the aromatic molecule set
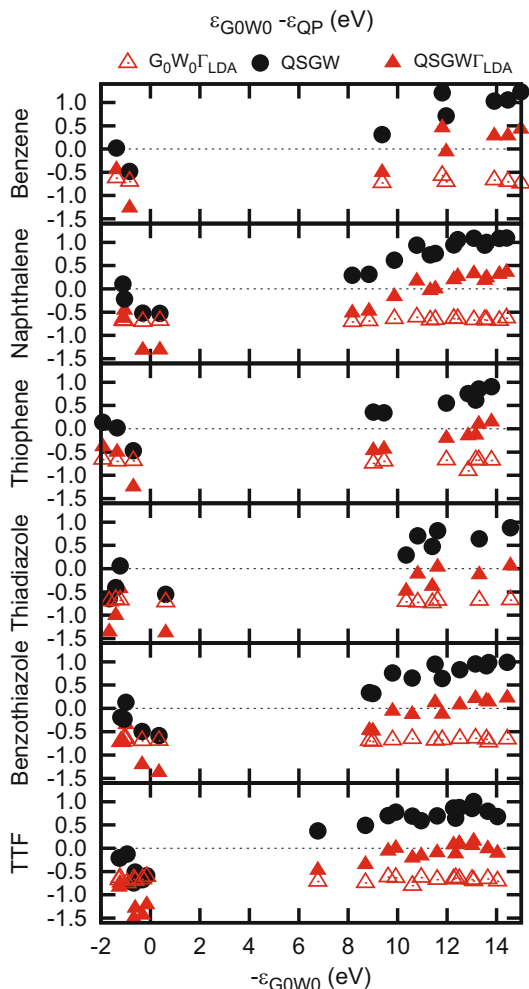
by ∼0.7 eV for $G_0W_0\Gamma_{\mathrm{LDA}}$ compared to $G_0W_0$. As a result, going from $G_0W_0$ to $G_0W_0\Gamma_{\mathrm{LDA}}$ barely affects the fundamental gap.

Figure 3 also shows that quasiparticle self-consistency makes the occupied orbitals more bound compared to the perturbative calculations. The first IP is slightly shifted upward, and deeper orbitals exhibit a greater shift. In contrast, EAs move in the opposite direction from IPs and become more negative compared to perturbative results. Quasiparticle self-consistency therefore increases the fundamental gap by 0.8–0.9 eV.

The two trends of the near-constant shift of eigenvalues after applying $\Gamma_{\mathrm{LDA}}$ and the opening of the $GW$ gap upon self-consistency agree with other benchmark studies for both confined and extended systems. In addition, these trends are also observed when both the LDA vertex and self-consistency are applied in combination through QS$GW\Gamma_{\mathrm{LDA}}$. The IPs predicted by QS$GW\Gamma_{\mathrm{LDA}}$ remain nearly a constant shift below those predicted by standard QS$GW$ with an energy difference ∼0.8 eV, and the EAs computed via QS$GW\Gamma_{\mathrm{LDA}}$ are on average ∼0.8 eV lower than those computed with QS$GW$. Again, the fundamental gap, predicted with $\Gamma_{\mathrm{LDA}}$, is essentially unchanged from the gap of conventional QS$GW$.

For the IPs, these benchmarks show a cancellation between the effects of vertex corrections and $GW$ self-consistency, as reported in earlier work (Shirley 1996; Ummels et al. 1998). For EAs, on the other hand, we see that the vertex and self-consistency effects are still roughly the same magnitude, but they shift in the same direction from $G_0W_0$ predictions, such that QS$GW\Gamma_{\mathrm{LDA}}$ compounds their effects instead of canceling them out. The comparable magnitudes of self-consistency and vertex corrections are likely coincidental, although it may be worth studying whether such cancellations for IP predictions are inherent to the physical properties of aromatic or $sp$-bonded molecules.

In benchmarks relative to experiment, we find that despite the greater effort in $\Gamma_{\mathrm{LDA}}$ and self-consistent computations, $G_0W_0$ predictions are closest to measured IPs. Experimental measurements exist for IPs up to 15 eV in most of the molecules, and $G_0W_0$ energies all lie within 0.6 eV of measured values with a mean absolute error less than 0.3 eV (Figs. 1 and 4). Predictions of IPs from $G_0W_0\Gamma_{\mathrm{LDA}}$ are too low, consistent with a benchmark of single atoms (Morris et al. 2007). The quasiparticle self-consistent results reverse the trend: an increase of the QS$GW$ IPs from perturbative values results in decreased accuracy, but the increase in QS$GW\Gamma_{\mathrm{LDA}}$ IPs improves agreement with experiment – although not to the extent of improving on the $G_0W_0$ values.

Experimental comparisons for EAs are only available for benzene and naphthalene, which are reported in Table 1 together with our $GW$ predictions. In contrast to the IPs, it is $G_0W_0\Gamma_{\mathrm{LDA}}$ and QS$GW$ that best predict the EAs. Interestingly, EAs from $G_0W_0\Gamma_{\mathrm{LDA}}$ are in good agreement with coupled-cluster [CCSD(T)] calculations, although both are too negative relative to experiment. It is unclear whether these differences are due to the difficulties in experimental measurements or an inaccurate representation of the unoccupied orbitals in the calculations. For instance, the limited basis set used in the higher-order corrections of CCSD(T) may

**Fig. 4** Error for the first IP (top) and the mean absolute error of orbitals with IPs up to 15 eV (bottom), relative to experiment for each molecule
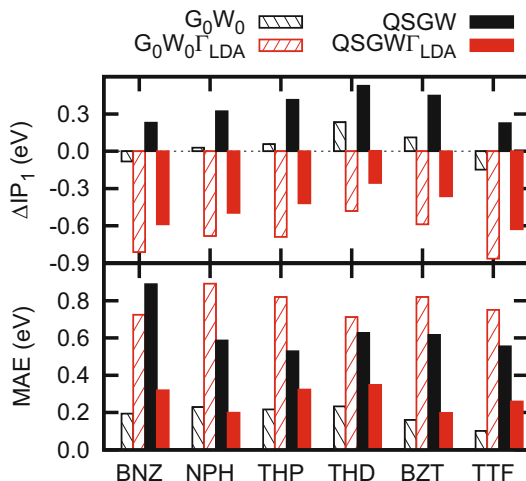


**Table 1** First EA in eV with comparison to electron transmission spectroscopy (ETS) measurements (Burrow et al. 1987) and CCSD(T) (Hajgató et al. 2008) calculations

| Molecule | Orbital | ETS | CCSD(T) | $G_0W_0$ | $G_0W_0\Gamma_{LDA}$ | QS$GW$ | QS$GW\Gamma_{LDA}$ |
|---|---|---|---|---|---|---|---|
| Benzene | $1e_{2u}$ ($\pi^*$) | −1.12 | −1.526 | −0.84 | −1.55 | −1.33 | −2.12 |
| Naphthalene | $2b_{1g}$ ($\pi^*$) | −0.19 | −0.477 | 0.38 | −0.30 | −0.15 | −0.93 |

numerically bind the wave functions, and similarly, our perturbative calculations use a DFT-LDA electronic structure, a theory which over-binds the unoccupied orbitals.

When we take the difference of predicted IP and EA for each molecule, we find that the fundamental gaps of benzene and naphthalene computed using QS$GW$ (11.01 and 8.61 eV, respectively) and QS$GW\Gamma_{LDA}$ (10.98 and 8.57 eV) are nearly identical to fundamental gaps determined from experimental IPs and CCSD(T) EAs (10.98 and 8.62 eV). However, we also see that no single variant of $GW$ studied here is best for the absolute IPs and EAs. For perturbative calculations, $G_0W_0$ is more accurate for IPs, and $G_0W_0\Gamma_{LDA}$ is more accurate for EAs, while among self-consistent calculations (which reduce or eliminate starting-point dependence), QS$GW\Gamma_{LDA}$ is more accurate for IPs, and QS$GW$ is more accurate for EAs. This example simulating real molecules confirms an earlier prediction derived from model systems: a two-point DFT-derived vertex can alleviate self-screening errors felt by occupied orbitals, but only a three-point vertex can be expected to treat both occupied and unoccupied orbitals accurately (Romaniello et al. 2009).

Concerning self-consistency, these results contrast with multiple benchmarks which show improved spectral properties for molecules after eigenvalue, QS, or fully self-consistent $GW$ (Dahlen and van Leeuwen (2005), Stan et al. (2006), and Stan et al. (2009) and some references in Sect. 1). These computations are performed using various numerical implementations. Therefore, for self-consistent

*GW*, numerical considerations such as the choice of a quasiparticle basis for self-consistency and the basis set chosen to represent wave functions must be better understood before a consensus can be reached on the theoretical accuracy of self-consistent *GW* for molecules.

Eigenvalue self-consistency and the LDA vertex are also used to compute *GW* quasiparticle energies of Cu, Ag, Zn, and Cd atoms in neutral and charged configurations, with comparison to experiment in Fig. 5. Similar to QS*GW* with the aromatic molecules, ev*GW* widens the fundamental gap, although not all quasiparticle energies are affected to the same degree. The ev*GW* energies for the HOMOs with *s* character ($d^{10}s$ and $d^{10}s^2$ valence) remain fairly similar to perturbative values. However, the energies of the LUMOs (all of which have *s* character) increase, and energies of the HOMO with *d* character ($d^{10}$ valence) decrease. The resulting *d* quasiparticle energies from ev*GW* are less accurate than in $G_0W_0$, and there is no systematic improvement for *s* or *p* states. Again, the



**Fig. 5** Error of *GW* quasiparticle energies for Cu, Ag, Zn, and Cd atoms and ions, relative to experiment for various flavors of *GW*. The reference valence electron configurations are given in bold on the right
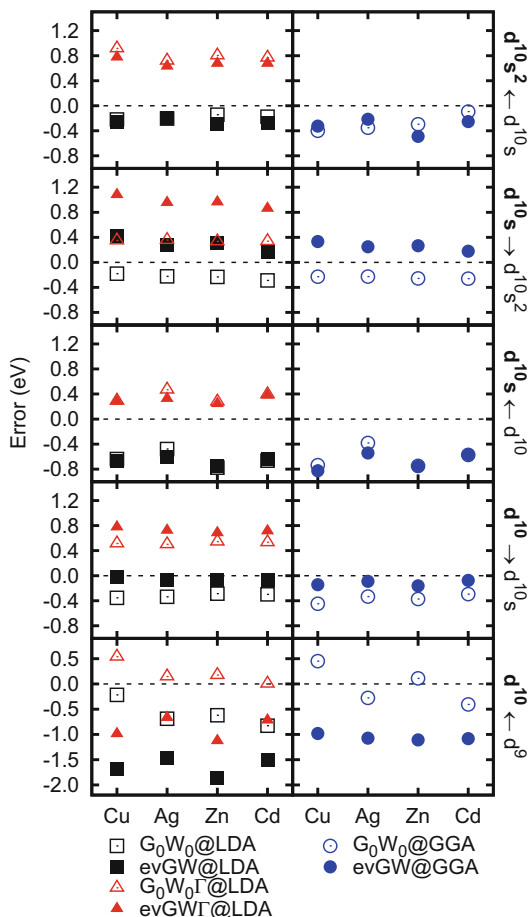
worsened quasiparticle energies differ from the majority of self-consistent $GW$ benchmarks (see list of benchmarks in Sect. 1), but similar results were observed in earlier studies of azabenzenes and small transition metal molecules (Marom et al. 2012; Körbel et al. 2014).

A more unusual result is observed for the LDA-vertex-corrected $GW$ benchmarks. Unlike the aromatic molecule test set (and other earlier studies), the energy change associated with the vertex correction is not constant but rather ranges between 0.6 and 1.1 eV. The variation is not due to differences in angular momentum character – indeed, the energy differences for $d$-states are $\sim$0.8 eV, in the middle of the observed range. However, the LDA-derived vertex function's sensitivity to the local wave function amplitude is showcased by the character of orbitals on single atoms, which range from localized to diffuse. We observe that the LDA vertex affects the quasiparticle energies most dramatically for the HOMO corresponding to the $d^{10}s \leftarrow d^{10}$ and the $d^{10}s^2 \leftarrow d^{10}s$ excitations; the quasiparticle wave functions used in computing these states are overly delocalized due to the use of the LDA exchange-correlation functional. In contrast, the LDA vertex changes the quasiparticle energies the least for the LUMO corresponding to the $d^{10}s \rightarrow d^{10}s^2$ excitation; this wave function is overly localized by the LDA exchange-correlation functional.

Despite the increased versatility of the LDA vertex in this context, the inclusion of $\Gamma_{\mathrm{LDA}}$ still does not improve agreement with experiment, and $G_0W_0$ remains the best predictor of ionization energies for these single atoms. Eigenvalue self-consistency and the LDA vertex correction together also do not exhibit any fortuitous cancellation of effects for this test set.

These benchmarks show that, although self-consistent $GW$ solutions have certain theoretical advantages, such as starting-point independence, fulfillment of energy and momentum conservation laws, and consistent values for observables when using different partitioning functions (Baym 1962), they do not necessarily perform well for the spectral properties of confined systems. In fact, for fully self-consistent $GW$, inaccuracies in computed values are unavoidable as long as the vertex approximation remains. This is the same as for extended (bulk) systems, where ground-state properties are in general accurately represented by self-consistent $GW$, but band energies can be less accurate than perturbative calculations, and valence bandwidths are too large (Ku and Eguiluz 2002; Kutepov et al. 2009).

## 4.4   $GW$-BSE

When computing two-particle interactions within the $GW$-BSE framework, the initial errors in the quasiparticles predicted via $GW$ (and DFT) may be propagated, amplified, or canceled out. Therefore, the level of theory best for $GW$-BSE must be confirmed in benchmarks separate from the $GW$ approximation alone.

We begin with a quick check of numerical convergence in the top row of Fig. 6, which contains TDDFT (with LDA exchange-correlation) spectra for the aromatic molecules in our test set computed in two different-sized simulation cells. For
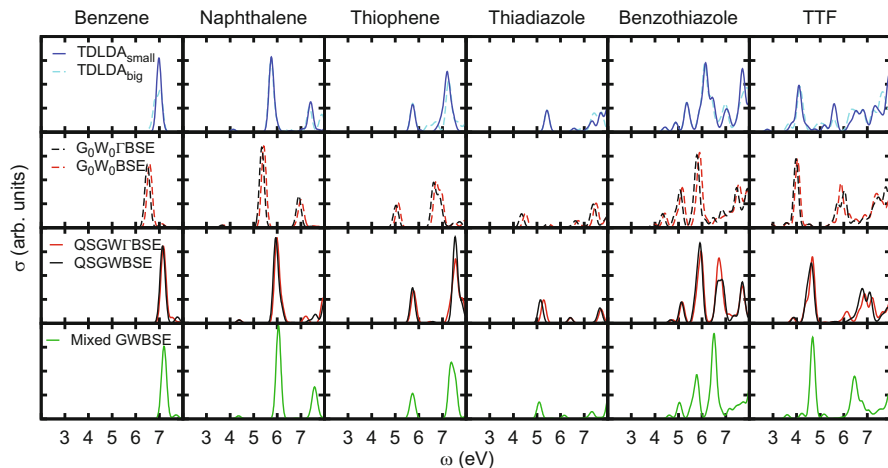
**Fig. 6** The absorption cross section for each of the molecules as predicted by various levels of theory (with Gaussian broadening of 0.1 eV). The TDLDA$_{big}$ calculations are performed in simulation cells with radii of 20 bohr, while TDLDA$_{small}$ simulation cells have radii that are either 12 bohr (for benzene, thiophene, and 1,2,5-thiadiazole) or 14 bohr (for naphthalene, benzothiazole, and TTF). $GW$-BSE calculations are performed in the smaller simulation cells

both cell sizes, the lowest-energy excitations are in agreement; deviations are only observed for the higher-energy excitations that involve transitions to regions farther from the atomic cores. This suggests that the simulation cell size is numerically converged for the lower-energy, localized excitations. Figure 6 also illustrates the trends in the $GW$-BSE energies (in the diagonal approximation) occurring due to self-consistency and the vertex function. The spectra from perturbative $GW$-BSE ($G_0W_0$-BSE and $G_0W_0\Gamma_{LDA}$-BSE) are all redshifted compared to the other spectra.

The trends due to self-consistency and the vertex approximation are also reflected in the detailed comparison of $GW$-BSE predictions of low-lying valence (localized) singlet excitations, in the top left panel of Fig. 7. All calculations are presented as the difference from mixed $GW$-BSE predictions, which we explain later in this section. In these benchmarks, the choice of vertex only minimally changes excitation energies. This is unsurprising, since as mentioned in the previous section, the energy differences between quasiparticles (both holes and quasielectrons) upon adding the LDA vertex essentially amount to a rigid shift, while energy differences between quasiparticle levels remain unchanged. The small differences in transition energies upon inclusion of $\Gamma_{LDA}$ – increasing an average of 0.09 eV for perturbative $GW$-BSE and 0.02 eV for QS$GW$-BSE – are attributed to the inclusion of $\Gamma_{LDA}$ in the BSE kernel. In contrast, self-consistency in $GW$-BSE has a large effect on excitation energies. QS$GW$-BSE has an average 0.91 eV increase of singlet excitation energies over $G_0W_0$-BSE, and QS$GW\Gamma_{LDA}$-BSE has an average 0.84 eV increase compared to $G_0W_0\Gamma_{LDA}$-BSE.
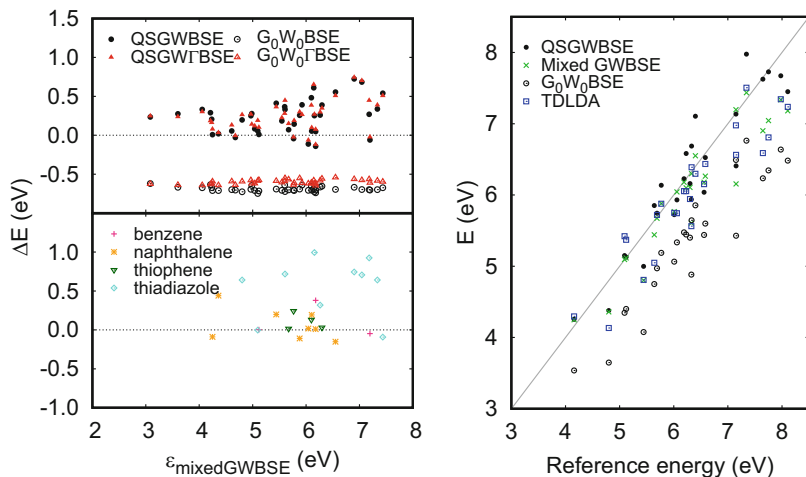
**Fig. 7** Left: with mixed $GW$-BSE predictions as reference, the top panel plots the difference between vertical valence excitation energies predicted by several variants of $GW$-BSE; the bottom panel is the difference between mixed $GW$-BSE and literature values computed using quantum chemistry methods (see text for references). Right: the difference between TDDFT and $GW$-BSE predictions and literature "reference" values

A comparison of $GW$-BSE and TDDFT predictions relative to best previous theoretical results for the aromatic molecules is shown in the right panel of Fig. 7. For benzene, naphthalene, and thiophene, the best previous theoretical values were computed including contributions from singles, doubles, and triples excitations (Falden et al. 2009; Fliegl and Sundholm 2014; Stenrup 2012). The mean absolute difference across these three molecules is 0.78 eV for $G_0W_0$-BSE, 0.68 eV for $G_0W_0\Gamma_{LDA}$, 0.25 eV for QS$GW$-BSE, 0.24 eV for QS$GW\Gamma_{LDA}$-BSE, and 0.25 eV for TDDFT. For 1,2,5-thiadiazole, the cited calculation only includes singles and doubles excitations (Palmer 2008), and the mean absolute difference is larger at 1.26 eV for $G_0W_0$-BSE, 1.14 eV for $G_0W_0\Gamma_{LDA}$, 0.33 eV for QS$GW$-BSE, 0.29 for $GW\Gamma_{LDA}$-BSE, and 0.61 for TDDFT. We observe that BSE calculations that build on perturbative $GW$ calculations produce singlet excitation energies that are too small by 0.7–0.8 eV compared to previous theoretical computations of benzene, naphthalene, and thiophene. On the other hand, QS$GW$-BSE and QS$GW\Gamma_{LDA}$-BSE have mean signed differences that are negative for benzene but positive for naphthalene and thiophene.

In Figs. 6 and 7, we also mention a "mixed $GW$-BSE" method. This is where we apply the BSE to a mixed set of $GW$ quasiparticles: occupied orbitals are associated with $G_0W_0$ quasiparticle energies, unoccupied orbitals have $G_0W_0\Gamma_{LDA}$ quasiparticle energies, and the screened interaction is computed using LDA quantities without vertex contributions. The quasiparticle wave functions in this case are given by the DFT wave functions. The mixed $GW$-BSE calculation is motivated by the observation in the previous section that $G_0W_0$ energies have the best agreement with

experimental values for IPs, while $G_0W_0\Gamma_{LDA}$ energies are better for EAs. Through this construction, mixed $GW$-BSE has an optical gap larger than perturbative $GW$-BSE calculations, which better matches reference values. For this set of molecules, the mean signed difference between mixed $GW$-BSE and previous theoretical results is approximately $-0.1$ eV, but we emphasize that this observed cancellation of effects is not necessarily transferable across different physical systems.

Also in Figs. 6 and 7, TDDFT predictions for localized valence excitations of $sp$-bonded molecules are shown to be in good agreement with higher-level quantum chemistry calculations. However, TDDFT with the LDA functional cannot properly simulate excitations such as charge-transfer and Rydberg excitations, where long-range interactions become important (Tozer and Handy 1998; Casida et al. 1998). In such cases, TDDFT functionals that better account for exchange and correlation interactions must be used to produce good optical spectra for molecules (Refaely-Abramson et al. 2011; Leang et al. 2012; Laurent and Jacquemin 2013). This reliance on specific exchange-correlation functionals contrasts with the self-consistent $GW$-BSE framework, which – assuming an adequate starting point – treats all excitations on equal footing.

We now turn to predictions from variants of the $GW$-BSE method for optical excitations in the Group IB and IIB atoms and compare them with spin-orbit-averaged experimental absorption energies and TDDFT eigenvalues. We focus on low-lying, spin-conserving excitations promoting electrons from the HOMO to the unoccupied valence $s$ or $p$ shells. Here, we also move beyond the diagonal approximation (Eq. 18), which was used for all $GW$-BSE energies computed for the above aromatic molecules, and determine BSE eigenvalues while also accounting for off-diagonal terms of the $GW$ self-energy. Figure 8 and Table 2 summarize the $GW$-BSE error relative to experiment, with the rectangles in Fig. 8 indicating the ranges from minimum to maximum error across the Cu, Ag, Zn, and Cd test set.

In the diagonal approximation, perturbative $GW$-BSE eigenvalues underestimate absorption energies, with improvements in accuracy for self-consistent $GW$-BSE, a trend consistent with other recent $GW$-BSE benchmarks of confined systems. The inclusion of the LDA vertex increases the predicted energies of optical excitations; however, without self-consistency (BSE@$G_0W_0\Gamma_{LDA}$@LDA), this still underestimates excitation energies. This is the same trend as observed for the aromatic molecules above, where perturbative $GW$-BSE excitation energies are too small, and some amount of self-consistency must be incorporated for better agreement with reference energies.

With the inclusion of off-diagonal terms, however, the predicted absorption energies increase for all variants of $GW$-BSE. This effect occurs because the low-lying LDA (and GGA) unoccupied orbitals are overly localized, and off-diagonal terms help delocalize the quasiparticles. With this energy increase, eigenvalue self-consistency no longer improves accuracy for all calculations. BSE@ev$GW$@LDA deteriorates in accuracy compared to BSE@$G_0W_0$@LDA, but BSE@ev$GW$@GGA is the most accurate variant of conventional $GW$ tested, with errors no more than $0.5$ eV. We also find that the computationally efficient BSE@$G_0W_0\Gamma_{LDA}$@LDA provides excellent agreement with experiment that is

**Fig. 8** Error of $GW$-BSE predictions relative to experimental absorption energies from a $d^{10}$ (first four sets of bars), $d^{10}s$ (fifth set), or $d^{10}s^2$ (sixth set) electron configuration to the configuration listed along the horizontal axis. Rectangles matching the legend indicate the error range across the Cu, Ag, Zn, and Cd test set, with fainter colored bars providing a guide for the eye
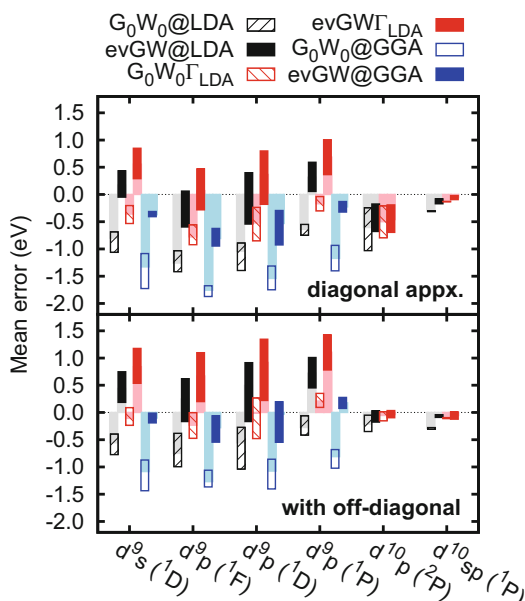


**Table 2** Mean error of $GW$-BSE eigenvalues (including off-diagonal terms) associated with excitations from the HOMO to the lowest empty $s$ or $p$ quasiparticle, relative to experimental energies

| | | BSE@ | | | | BSE@ | |
|---|---|---|---|---|---|---|---|
| | TDDFT | $G_0W_0$ | evGW | $G_0W_0\Gamma_{LDA}$ | evGW$\Gamma_{LDA}$ | $G_0W_0$ | evGW |
| Transition | @LDA | @LDA | | | | @GGA | |
| $d^{10}s^2 \to d^{10}sp$ ($^1$P) | −0.03 | −0.29 | −0.07 | −0.10 | −0.06 | – | – |
| $d^{10}s \to d^{10}p$ ($^2$P) | 0.51 | −0.21 | −0.09 | −0.05 | −0.04 | – | – |
| $d^{10} \to d^9p$ ($^1$P) | −1.48 | −0.28 | 0.72 | 0.21 | 1.11 | −0.81 | 0.18 |
| $d^{10} \to d^9p$ ($^1$D) | −1.90 | −0.56 | 0.50 | −0.01 | 0.92 | −1.08 | −0.03 |
| $d^{10} \to d^9p$ ($^1$F) | −1.95 | −0.75 | 0.25 | −0.23 | 0.67 | −1.27 | −0.28 |
| $d^{10} \to d^9s$ ($^1$D) | −1.54 | −0.56 | 0.44 | −0.11 | 0.84 | −1.08 | −0.10 |

comparable to BSE@ev$GW$@GGA. While further study is needed, at least for the Group IB and IIB test set, DFT-derived vertex corrections combined with off-diagonal terms allow an accurate calculation of BSE energies from less exact $GW$ quasiparticles.

## 5    Summary

Computational advances now make it feasible – if still expensive – to perform $GW$ calculations at varying levels of theory, and researchers have begun to understand the intricate approximations used to model the excited state from first principles.

In this chapter, we discussed many-body perturbation theory methods that allow the simulation of the excited state in confined systems. We overviewed an efficient implementation of $GW$, BSE, and TDDFT that uses a transition-space basis to describe excitations and presented benchmarks of excitations in atoms and molecules, comparing approximations taken in theory and implementation. Since the accuracy of variants of $GW$ and $GW$-BSE can be compared only after knowing the numerical accuracy for context, we discussed the impact of finite basis sets, including methods to estimate quasiparticle energies at the complete basis set limit.

Using test sets of aromatic molecules and transition metal atoms and ions, we examined the impact of self-consistency within $GW$ as well as the vertex approximation. In our benchmarks involving self-consistency, we found worsened accuracy compared to $G_0W_0$ results but caution that the performance of these and other benchmarks may still be influenced by numerical approximations, not just theory. For the LDA-derived vertex, the results presented here generally agreed with past work, where the vertex shifted both occupied and unoccupied states to lower energies compared to conventional $GW$. However, while all past results reported shifts of a near-constant magnitude, the LDA vertex applied to transition metal atoms resulted in shifts whose magnitude varied according to the diffusivity of associated orbitals. We also addressed the off-diagonal terms of the self-energy. These did not necessarily improve accuracy in $GW$ predictions (such as in QS$GW$) but did contribute to improved descriptions of neutral excitations within the perturbative $GW$-BSE framework.

# References

Anglade PM, Gonze X (2008) Preconditioning of self-consistent-field cycles in density-functional theory: the extrapolar method. Phys Rev B 78(4):045126. https://link.aps.org/doi/10.1103/PhysRevB.78.045126

Atalla V, Yoon M, Caruso F, Rinke P, Scheffler M (2013) Hybrid density functional theory meets quasiparticle calculations: a consistent electronic structure approach. Phys Rev B 88(16):165122. http://link.aps.org/doi/10.1103/PhysRevB.88.165122

Baumeier B, Andrienko D, Ma Y, Rohlfing M (2012) Excited states of Dicyanovinyl-substituted Oligothiophenes from many-body Green's functions theory. J Chem Theory Comput 8(3):997–1002. https://doi.org/10.1021/ct2008999

Baym G (1962) Self-consistent approximations in many-body systems. Phys Rev 127(4):1391–1401. http://link.aps.org/doi/10.1103/PhysRev.127.1391

Berger JA, Reining L, Sottile F (2010) *Ab initio* calculations of electronic excitations: collapsing spectral sums. Phys Rev B 82(4):041103. http://link.aps.org/doi/10.1103/PhysRevB.82.041103

Blase X, Attaccalite C (2011) Charge-transfer excitations in molecular donor-acceptor complexes within the many-body Bethe-Salpeter approach. Appl Phys Lett 99(17):171909. http://scitation.aip.org/content/aip/journal/apl/99/17/10.1063/1.3655352

Blase X, Attaccalite C, Olevano V (2011) First-principles GW calculations for fullerenes, por-
phyrins, phthalocyanine, and other molecules of interest for organic photovoltaic applications.
Phys Rev B 83(11):115103. http://link.aps.org/doi/10.1103/PhysRevB.83.115103

Blase X, Boulanger P, Bruneval F, Fernandez-Serra M, Duchemin I (2016) GW and Bethe-Salpeter
study of small water clusters. J Chem Phys 144(3):034109. http://scitation.aip.org/content/aip/
journal/jcp/144/3/10.1063/1.4940139

Boulanger P, Jacquemin D, Duchemin I, Blase X (2014) Fast and accurate electronic excitations in
cyanines with the many-body Bethe–Salpeter approach. J Chem Theory Comput 10(3):1212–
1218. https://doi.org/10.1021/ct401101u

Bruneval F (2012) Ionization energy of atoms obtained from GW self-energy or from random phase
approximation total energies. J Chem Phys 136(19):194107. http://scitation.aip.org/content/aip/
journal/jcp/136/19/10.1063/1.4718428

Bruneval F, Gonze X (2008) Accurate GW self-energies in a plane-wave basis using only a few
empty states: towards large systems. Phys Rev B 78(8):085125. http://link.aps.org/doi/10.1103/
PhysRevB.78.085125

Bruneval F, Marques MAL (2013) Benchmarking the starting points of the GW approximation for
molecules. J Chem Theory Comput 9(1):324–329. https://doi.org/10.1021/ct300835h

Bruneval F, Sottile F, Olevano V, Del Sole R, Reining L (2005) Many-body perturbation
theory using the density-functional concept: beyond the GW approximation. Phys Rev Lett
94(18):186402. http://link.aps.org/doi/10.1103/PhysRevLett.94.186402

Bruneval F, Hamed SM, Neaton JB (2015) A systematic benchmark of the ab initio Bethe-Salpeter
equation approach for low-lying optical excitations of small organic molecules. J Chem Phys
142(24):244101. http://scitation.aip.org/content/aip/journal/jcp/142/24/10.1063/1.4922489

Bruneval F, Rangel T, Hamed SM, Shao M, Yang C, Neaton JB (2016) MolGW 1: many-
body perturbation theory software for atoms, molecules, and clusters. Comput Phys Com-
mun 208:149–161. https://doi.org/10.1016/j.cpc.2016.06.019, http://www.sciencedirect.com/
science/article/pii/S0010465516301990

Burrow PD, Michejda JA, Jordan KD (1987) Electron transmission study of the temporary negative
ion states of selected Benzenoid and conjugated aromatic hydrocarbons. J Chem Phys 86(1):9–
24. http://scitation.aip.org/content/aip/journal/jcp/86/1/10.1063/1.452598

Caruso F, Rinke P, Ren X, Rubio A, Scheffler M (2013) Self-consistent GW: all-electron
implementation with localized basis functions. Phys Rev B 88(7):075105. http://link.aps.org/
doi/10.1103/PhysRevB.88.075105

Casida ME (2009) Time-dependent density-functional theory for molecules and molecular solids.
J Mol Struc Theochem 914(1–3):3–18. https://doi.org/10.1016/j.theochem.2009.08.018, http://
www.sciencedirect.com/science/article/pii/S0166128009005363

Casida ME, Jamorski C, Casida KC, Salahub DR (1998) Molecular excitation energies to high-
lying bound states from time-dependent density-functional response theory: characterization
and correction of the time-dependent local density approximation ionization threshold. J Chem
Phys 108(11):4439. https://doi.org/10.1063/1.475855

Dahlen NE, van Leeuwen R (2005) Self-consistent solution of the Dyson equation for atoms and
molecules within a conserving approximation. J Chem Phys 122(16):164102. http://scitation.
aip.org/content/aip/journal/jcp/122/16/10.1063/1.1884965

Del Sole R, Reining L, Godby RW (1994) GW$\Gamma$ approximation for electron self-energies in
semiconductors and insulators. Phys Rev B 49(12):8024–8028. http://link.aps.org/doi/10.1103/
PhysRevB.49.8024

Deslippe J, Samsonidze G, Strubbe DA, Jain M, Cohen ML, Louie SG (2012) BerkeleyGW:
a massively parallel computer package for the calculation of the quasiparticle and optical
properties of materials and nanostructures. Comput Phys Commun 183(6):1269–1289. http://
www.sciencedirect.com/science/article/pii/S0010465511003912

Deslippe J, Samsonidze G, Jain M, Cohen ML, Louie SG (2013) Coulomb-hole summations and
energies for GW calculations with limited number of empty orbitals: a modified static remainder
approach. Phys Rev B 87(16):165124. http://link.aps.org/doi/10.1103/PhysRevB.87.165124

Faber C, Attaccalite C, Olevano V, Runge E, Blase X (2011) First-principles *GW* calculations for DNA and RNA nucleobases. Phys Rev B 83(11):115123. http://link.aps.org/doi/10.1103/PhysRevB.83.115123

Falden HH, Falster-Hansen KR, Bak KL, Rettrup S, Sauer SPA (2009) Benchmarking second order methods for the calculation of vertical electronic excitation energies: valence and Rydberg states in polycyclic aromatic hydrocarbons†. J Phys Chem A 113(43):11995–12012. https://doi.org/10.1021/jp9037123

Faleev SV, van Schilfgaarde M, Kotani T (2004) All-electron self-consistent *GW* approximation: application to Si, MnO, and NiO. Phys Rev Lett 93(12):126406. http://link.aps.org/doi/10.1103/PhysRevLett.93.126406

Fliegl H, Sundholm D (2014) Coupled-cluster calculations of the lowest 0–0 bands of the electronic excitation spectrum of naphthalene. Phys Chem Chem Phys 16(21):9859–9865. https://doi.org/10.1039/C3CP54421D, http://pubs.rsc.org/en/content/articlelanding/2014/cp/c3cp54421d

Gao W, Xia W, Gao X, Zhang P (2016) Speeding up GW calculations to meet the challenge of large scale Quasiparticle predictions. Sci Rep 6:36849. https://doi.org/10.1038/srep36849, http://www.nature.com/srep/2016/161111/srep36849/full/srep36849.html

Godby RW, Schlüter M, Sham LJ (1988) Self-energy operators and exchange-correlation potentials in semiconductors. Phys Rev B 37(17):10159–10175. http://link.aps.org/doi/10.1103/PhysRevB.37.10159

Grüneis A, Kresse G, Hinuma Y, Oba F (2014) Ionization potentials of solids: the importance of vertex corrections. Phys Rev Lett 112(9):096401. http://link.aps.org/doi/10.1103/PhysRevLett.112.096401

Gulans A (2014) Towards numerically accurate many-body perturbation theory: short-range correlation effects. J Chem Phys 141(16):164127. http://scitation.aip.org/content/aip/journal/jcp/141/16/10.1063/1.4900447

Hajgató B, Deleuze MS, Tozer DJ, De Proft F (2008) A benchmark theoretical study of the electron affinities of benzene and linear acenes. J Chem Phys 129(8):084308. http://scitation.aip.org/content/aip/journal/jcp/129/8/10.1063/1.2967182

Hedin L (1965) New method for calculating the one-particle green's function with application to the electron-gas problem. Phys Rev 139(3A):A796–A823. http://link.aps.org/doi/10.1103/PhysRev.139.A796

Hirose D, Noguchi Y, Sugino O (2015) All-electron *GW*+Bethe-Salpeter calculations on small molecules. Phys Rev B 91(20):205111. http://link.aps.org/doi/10.1103/PhysRevB.91.205111

Hung L, Baishya K, Öğüt S (2014) First-principles real-space study of electronic and optical excitations in rutile $TiO_2$ nanocrystals. Phys Rev B 90(16):165424. http://link.aps.org/doi/10.1103/PhysRevB.90.165424

Hung L, da Jornada FH, Souto-Casares J, Chelikowsky JR, Louie SG, Öğüt S (2016) Excitation spectra of aromatic molecules within a real-space *GW*-BSE formalism: role of self-consistency and vertex corrections. Phys Rev B 94(8):085125. http://link.aps.org/doi/10.1103/PhysRevB.94.085125

Hung L, Bruneval F, Baishya K, Öğüt S (2017) Benchmarking the GW approximation and Bethe–Salpeter equation for groups IB and IIB atoms and monoxides. J Chem Theory Comput 13(5):2135–2146. https://doi.org/10.1021/acs.jctc.7b00123

Hybertsen MS, Louie SG (1986) Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. Phys Rev B 34(8):5390–5413. http://link.aps.org/doi/10.1103/PhysRevB.34.5390

Jacquemin D, Duchemin I, Blase X (2015) Benchmarking the Bethe–Salpeter formalism on a standard organic molecular set. J Chem Theory Comput 11(7):3290–3304. https://doi.org/10.1021/acs.jctc.5b00304

Jiang H, Blaha P (2016) *GW* with linearized augmented plane waves extended by high-energy local orbitals. Phys Rev B 93(11):115203. http://link.aps.org/doi/10.1103/PhysRevB.93.115203

Kang W, Hybertsen MS (2010) Enhanced static approximation to the electron self-energy operator for efficient calculation of quasiparticle energies. Phys Rev B 82(19):195108. http://link.aps.org/doi/10.1103/PhysRevB.82.195108

Kaplan F, Weigend F, Evers F, van Setten MJ (2015) Off-diagonal self-energy terms and partially self-consistency in GW calculations for single molecules: efficient implementation and quantitative effects on ionization potentials. J Chem Theory Comput 11(11):5152–5160. https://doi.org/10.1021/acs.jctc.5b00394

Kaplan F, Harding ME, Seiler C, Weigend F, Evers F, van Setten MJ (2016) Quasi-particle self-consistent GW for molecules. J Chem Theory Comput 12(6):2528–2541. https://doi.org/10.1021/acs.jctc.5b01238

Ke SH (2011) All-electron *GW* methods implemented in molecular orbital space: ionization energy and electron affinity of conjugated molecules. Phys Rev B 84(20):205415. http://link.aps.org/doi/10.1103/PhysRevB.84.205415

Klimeš J, Kaltak M, Kresse G (2014) Predictive GW calculations using plane waves and pseudopotentials. Phys Rev B 90(7):075125. http://link.aps.org/doi/10.1103/PhysRevB.90.075125

Knight JW, Wang X, Gallandi L, Dolgounitcheva O, Ren X, Ortiz JV, Rinke P, Körzdörfer T, Marom N (2016) Accurate ionization potentials and electron affinities of acceptor molecules III: a benchmark of GW methods. J Chem Theory Comput 12(2):615–626. https://doi.org/10.1021/acs.jctc.5b00871

Körbel S, Boulanger P, Duchemin I, Blase X, Marques MAL, Botti S (2014) Benchmark many-body GW and Bethe–Salpeter calculations for small transition metal molecules. J Chem Theory Comput 10(9):3934–3943. https://doi.org/10.1021/ct5003658

Koval P, Foerster D, Sánchez-Portal D (2014) Fully self-consistent *GW* and quasiparticle self-consistent $GW$ for molecules. Phys Rev B 89(15):155417. http://link.aps.org/doi/10.1103/PhysRevB.89.155417

Krause K, Harding ME, Klopper W (2015) Coupled-cluster reference values for the GW27 and GW100 test sets for the assessment of GW methods. Mol Phys 113(13–14):1952–1960. http://www.tandfonline.com/doi/full/10.1080/00268976.2015.1025113

Ku W, Eguiluz AG (2002) Band-gap problem in semiconductors revisited: effects of core states and many-body self-consistency. Phys Rev Lett 89(12):126401. http://link.aps.org/doi/10.1103/PhysRevLett.89.126401

Kutepov A, Savrasov SY, Kotliar G (2009) Ground-state properties of simple elements from GW calculations. Phys Rev B 80(4):041103. http://link.aps.org/doi/10.1103/PhysRevB.80.041103

Laurent AD, Jacquemin D (2013) TD-DFT benchmarks: a review. Int J Quantum Chem 113(17):2019–2039. http://onlinelibrary.wiley.com/doi/10.1002/qua.24438/abstract

Leang SS, Zahariev F, Gordon MS (2012) Benchmarking the performance of time-dependent density functional methods. J Chem Phys 136(10):104101. http://scitation.aip.org/content/aip/journal/jcp/136/10/10.1063/1.3689445

Louie SG, Rubio A (2005) Quasiparticle and optical properties of solids and nanostructures: the GW-BSE approach. In: Handbook of materials modeling, Yip S (Ed.), Springer, Dordrecht, pp 215–240. https://link.springer.com/chapter/10.1007/978-1-4020-3286-8_12

Ma Y, Rohlfing M, Molteni C (2010) Modeling the excited states of biological chromophores within many-body green's function theory. J Chem Theory Comput 6(1):257–265. https://doi.org/10.1021/ct900528h

Maebashi H, Takada Y (2011) Analysis of exact vertex function for improving on the GWΓ scheme for first-principles calculation of electron self-energy. Phys Rev B 84(24):245134. http://link.aps.org/doi/10.1103/PhysRevB.84.245134

Marini A, Rubio A (2004) Electron linewidths of wide-gap insulators: excitonic effects in LiF. Phys Rev B 70(8):081103. http://link.aps.org/doi/10.1103/PhysRevB.70.081103

Marom N, Caruso F, Ren X, Hofmann OT, Körzdörfer T, Chelikowsky JR, Rubio A, Scheffler M, Rinke P (2012) Benchmark of GW methods for azabenzenes. Phys Rev B 86(24):245127. http://link.aps.org/doi/10.1103/PhysRevB.86.245127

Morris AJ, Stankovski M, Delaney KT, Rinke P, García-González P, Godby RW (2007) Vertex corrections in localized and extended systems. Phys Rev B 76(15):155106. http://link.aps.org/doi/10.1103/PhysRevB.76.155106

Onida G, Reining L, Rubio A (2002) Electronic excitations: density-functional versus many-body Green's-function approaches. Rev Mod Phys 74(2):601. http://link.aps.org/doi/10.1103/RevModPhys.74.601

Palmer MH (2008) The electronic states of 1,2,5-thiadiazole studied by VUV absorption spectroscopy and ab initio configuration interaction methods. Chem Phys 348(1–3):130–142. https://doi.org/10.1016/j.chemphys.2008.02.004, http://www.sciencedirect.com/science/article/pii/S0301010408001079

Pham TA, Nguyen HV, Rocca D, Galli G (2013) *GW* calculations using the spectral decomposition of the dielectric matrix: verification, validation, and comparison of methods. Phys Rev B 87:155148. http://journals.aps.org/prb/abstract/10.1103/PhysRevB.87.155148

Qian X, Umari P, Marzari N (2011) Photoelectron properties of DNA and RNA bases from many-body perturbation theory. Phys Rev B 84(7):075103. http://link.aps.org/doi/10.1103/PhysRevB.84.075103

Refaely-Abramson S, Baer R, Kronik L (2011) Fundamental and excitation gaps in molecules of relevance for organic photovoltaics from an optimally tuned range-separated hybrid functional. Phys Rev B 84(7):075144. http://link.aps.org/doi/10.1103/PhysRevB.84.075144

Ren X, Rinke P, Blum V, Wieferink J, Tkatchenko A, Sanfilippo A, Reuter K, Scheffler M (2012) Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. New J Phys 14(5):053020. https://doi.org/10.1088/1367-2630/14/5/053020, http://iopscience.iop.org/1367-2630/14/5/053020

Rocca D, Gebauer R, Saad Y, Baroni S (2008) Turbo charging time-dependent density-functional theory with Lanczos chains. J Chem Phys 128(15):154105. https://aip.scitation.org/doi/abs/10.1063/1.2899649

Rohlfing M, Louie SG (2000) Electron-hole excitations and optical spectra from first principles. Phys Rev B 62(8):4927–4944. http://link.aps.org/doi/10.1103/PhysRevB.62.4927

Romaniello P, Guyot S, Reining L (2009) The self-energy beyond GW: local and nonlocal vertex corrections. J Chem Phys 131(15):154111. http://scitation.aip.org/content/aip/journal/jcp/131/15/10.1063/1.3249965

Rostgaard C, Jacobsen KW, Thygesen KS (2010) Fully self-consistent GW calculations for molecules. Phys Rev B 81(8):085103. http://link.aps.org/doi/10.1103/PhysRevB.81.085103

Samsonidze G, Jain M, Deslippe J, Cohen ML, Louie SG (2011) Simple approximate physical orbitals for GW quasiparticle calculations. Phys Rev Lett 107(18):186404. http://link.aps.org/doi/10.1103/PhysRevLett.107.186404

Sharifzadeh S, Tamblyn I, Doak P, Darancet PT, Neaton JB (2012) Quantitative molecular orbital energies within a G0w0 approximation. Eur Phys J B 85(9):1–5. http://link.springer.com/article/10.1140/epjb/e2012-30206-0

Shirley EL (1996) Self-consistent *GW* and higher-order calculations of electron states in metals. Phys Rev B 54(11):7758–7764. http://link.aps.org/doi/10.1103/PhysRevB.54.7758

Stan A, Dahlen NE, van Leeuwen R (2006) Fully self-consistent GW calculations for atoms and molecules. Europhys Lett 76(2):298. https://doi.org/10.1209/epl/i2006-10266-6, http://iopscience.iop.org/0295-5075/76/2/298

Stan A, Dahlen NE, van Leeuwen R (2009) Levels of self-consistency in the GW approximation. J Chem Phys 130(11):114105. http://scitation.aip.org/content/aip/journal/jcp/130/11/10.1063/1.3089567

Stefanucci G, Pavlyukh Y, Uimonen AM, van Leeuwen R (2014) Diagrammatic expansion for positive spectral functions beyond *GW*: application to vertex corrections in the electron gas. Phys Rev B 90(11):115134. http://link.aps.org/doi/10.1103/PhysRevB.90.115134

Stenrup M (2012) Theoretical study of the radiationless deactivation mechanisms of photo-excited thiophene. Chem Phys 397:18–25. https://doi.org/10.1016/j.chemphys.2011.12.004, http://www.sciencedirect.com/science/article/pii/S0301010411005507

Strinati G (1988) Application of the green's functions method to the study of the optical properties of semiconductors. Riv Nuovo Cimento 11(12):1–86

Tiago ML, Chelikowsky JR (2006) Optical excitations in organic molecules, clusters, and defects studied by first-principles Green's function methods. Phys Rev B 73(20):205334. http://link.aps.org/doi/10.1103/PhysRevB.73.205334

Tozer DJ, Handy NC (1998) Improving virtual Kohn–Sham orbitals and eigenvalues: application to excitation energies and static polarizabilities. J Chem Phys 109(23):10180–10189. http://scitation.aip.org/content/aip/journal/jcp/109/23/10.1063/1.477711

Truhlar DG (1998) Basis-set extrapolation. Chem Phys Lett 294(1–3):45–48. https://doi.org/10.1016/S0009-2614(98)00866-5, http://www.sciencedirect.com/science/article/pii/S0009261498008665

Umari P, Stenuit G, Baroni S (2010) GW quasiparticle spectra from occupied states only. Phys Rev B 81(11):115104. http://link.aps.org/doi/10.1103/PhysRevB.81.115104

Ummels RTM, Bobbert PA, van Haeringen W (1998) First-order corrections to random-phase approximation *GW* calculations in silicon and diamond. Phys Rev B 57(19):11962–11973. http://link.aps.org/doi/10.1103/PhysRevB.57.11962

van Setten MJ, Caruso F, Sharifzadeh S, Ren X, Scheffler M, Liu F, Lischner J, Lin L, Deslippe JR, Louie SG, Yang C, Weigend F, Neaton JB, Evers F, Rinke P (2015) GW100: benchmarking G0w0 for molecular systems. J Chem Theory Comput 11(12):5665–5687. https://doi.org/10.1021/acs.jctc.5b00453

Wang LW (2015) Fully self-consistent solution of the Dyson equation using a plane-wave basis set. Phys Rev B 91(12):125135. http://link.aps.org/doi/10.1103/PhysRevB.91.125135

# Many-Body Calculations of Plasmon and Phonon Satellites in Angle-Resolved Photoelectron Spectra Using the Cumulant Expansion Approach

# 16

Fabio Caruso, Carla Verdi, and Feliciano Giustino

## Contents

### Abstract

The interaction of electrons with crystal lattice vibrations (phonons) and collective charge-density fluctuations (plasmons) influences profoundly the spectral properties of solids revealed by photoemission spectroscopy experiments. Photoemission satellites, for instance, are a prototypical example of quantum

F. Caruso
Institut für Physik and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin, Germany
e-mail: fabio.caruso@physik.hu-berlin.de

C. Verdi
Department of Materials, University of Oxford, Oxford, UK
e-mail: carla.verdi@materials.ox.ac.uk

F. Giustino (✉)
Department of Materials, University of Oxford, Oxford, UK

Department of Materials Science and Engineering, Cornell University, Ithaca, NY, USA
e-mail: feliciano.giustino@materials.ox.ac.uk

emergent behavior that may result from the strong coupling of electronic states to plasmons and phonons. The existence of these spectral features has been verified over energy scales spanning several orders of magnitude (from 50 meV to 15–20 eV) and for a broad class of compounds such as simple metals, semiconductors, and highly doped oxides. During the past few years, the cumulant expansion approach, alongside with the $GW$ approximation and the theory of electron-phonon and electron-plasmon coupling in solids, has evolved into a predictive and quantitatively accurate approach for the description of the spectral signatures of electron-boson coupling entirely from first principles, and it has thus become the state-of-the-art theoretical tool for the description of these phenomena. In this chapter we introduce the fundamental concepts needed to interpret plasmon and phonon satellites in photoelectron spectra, and we review recent progress on first-principles calculations of these features using the cumulant expansion method.

# 1 Introduction

The emergence of satellites in photoemission spectroscopy provides direct evidence of the electronic coupling to bosonic excitations in solids. Satellites are spectral features that reflect the simultaneous excitation of a hole and of a boson, and they are separated from the quasiparticle peak by a multiple of the boson energy. The origin of these features may be understood based on simple considerations on the energy scales involved in the photoemission process. When a photon with energy $\hbar\omega$ is absorbed by an electron with binding energy $\varepsilon_i$, if no boson modes are excited in the system, energy conservation requires the condition $\hbar\omega = \varepsilon_i + \Phi + E_{\text{kin}}$ to be satisfied, where $\Phi$ is the work function of the system and $E_{\text{kin}}$ is the kinetic energy of the photo-emitted electron. In photoemission, by measuring $E_{\text{kin}}$ and $\Phi$, the electron binding energy can thus be inferred. If, in addition to the creation of a hole, a fraction of the absorbed photon energy is transferred to the system in the form of bosonic modes, such as plasmons and phonons, the energy conservation condition is modified as follows: $E_{\text{kin}} = \hbar\omega - \varepsilon_i - nE_{\text{b}} - \Phi$, where $E_{\text{b}}$ is the energy of the boson and $n$ an integer. Since $\hbar\omega$ and $\Phi$ are constants, the kinetic energy distribution of the photo-emitted electrons will be peaked at the energies corresponding to (i) the binding energy of electrons $\varepsilon_i$ and (ii) the sum of the binding and boson energies $\varepsilon_i + nE_{\text{b}}$, and it may thus provide direct information regarding the coupling of electrons to bosonic modes in solids.

The presence of satellites in the photoemission spectra of solids was first predicted by a theoretical analysis of the spectral function of the homogeneous electron gas by Lundqvist (1967), Hedin and Lundqvist (1970), and Langreth (1970) and subsequently verified experimentally for the core electrons of simple metals (Baer and Busch 1973). Recently, the availability of energy resolutions of the order of 25–50 meV in angle-resolved photoelectron spectroscopy (ARPES) made it possible to observe new low-energy signatures of electron-boson coupling in experiments. In particular, high-resolution ARPES measurements of graphene

by Bostwick et al. (2010) have revealed plasmon-induced satellite structures with characteristic energies of the order of ~1 eV. More recently, polaronic satellites at energies of the order ~100 meV from the band edges have been observed in doped oxides, for example, by Moser et al. (2013) and Wang et al. (2016). At variance with valence-plasmon satellites, which typically appear at energies between 5 and 15 eV below the Fermi energy and have been known since the early days of photoemission spectroscopy, low-energy satellites are a manifestation of the coupling between low-energy bosonic modes and electronic carriers near the band edges. In addition to the formation of satellites, the coupling to bosons may lead to the emergence of photoemission kinks (Lanzara et al. 2001; Damascelli et al. 2003) and to a renormalization of energy levels (Logothetidis et al. 1992; Giustino et al. 2010; Poncé et al. 2015) and carrier lifetimes (Eiguren et al. 2002; Park et al. 2007).

In this chapter we will discuss the state-of-the-art techniques for the description of plasmon and polaron satellites and their application to the prediction and interpretation of photoemission spectroscopy experiments.

## 2   The Localized Electron Model

To illustrate how the interaction between electrons and bosons may lead to the emergence of satellites in photoemission spectra, we consider in the following the exactly solvable model of a "localized electron" in a solid interacting with a boson bath. The latter can be regarded as a set of phonons, plasmons, or any other bosonic excitations that may be approximately represented as a set of uncoupled harmonic oscillators. The localized electron is assumed to be dispersionless, that is, its energy $\varepsilon$ is independent of the crystal momentum, and its interaction with other electrons in the system is neglected. Instances in which the electron energy levels exhibit a weak dependence on momentum, and can thus be approximated as nondispersive, are, for example, core electrons in solids, localized impurity levels, and $4f$ electrons. On the other hand, electron-electron interactions are typically strong and non-negligible in three-dimensional solids, which poses limitations to the applicability of this model to real physical systems. This simplified model, however, is remarkably successful in describing the emergence of bosonic satellites in the spectral properties and is in good qualitative agreement with more advanced theories, whereby the electron-electron interaction is accounted for. This is demonstrated, for instance, by the generalization of the localized electron model reported by Langreth (1970).

The localized electron model is described by the following electron-boson coupling Hamiltonian:

$$\hat{H} = \hat{H}_e + \hat{H}_b + \hat{H}_{int} = \varepsilon \hat{c}^\dagger \hat{c} + \sum_{\mathbf{q}} \hbar \omega_{\mathbf{q}} \hat{b}_{\mathbf{q}}^\dagger \hat{b}_{\mathbf{q}} + \sum_{\mathbf{q}} g_{\mathbf{q}} \hat{c}^\dagger \hat{c} (\hat{b}_{\mathbf{q}} + \hat{b}_{-\mathbf{q}}^\dagger), \qquad (1)$$

where $\hat{c}^\dagger$ and $\hat{c}$ are fermionic creation and annihilation operators for the localized electron, respectively, which satisfy the ordinary anti-commutation relations.

Similarly, the operators $\hat{b}_{\mathbf{q}}^{\dagger}$ and $\hat{b}_{\mathbf{q}}$, respectively, create and annihilate a boson with energy $\hbar\omega_{\mathbf{q}}$ and momentum $\mathbf{q}$ and satisfy bosonic commutation relations. The absence of two-particle interaction terms in the Hamiltonian reflects the fact that both electron-electron and boson-boson interactions are neglected. The localized electron interacts with the boson bath with the coupling strength $g_{\mathbf{q}}$.

As we are primarily interested in the effects of the electron-boson interaction on the photoemission intensity, the relevant quantity that we want to compute is the electron spectral function:

$$A(\omega) = -\frac{1}{\pi}\mathrm{Im}\,G^{\mathrm{ret}}(\omega), \qquad (2)$$

with the single-particle retarded Green's function $G^{\mathrm{ret}}$ defined as:

$$G^{\mathrm{ret}}(t) = -i\,\langle\Psi_0|\,\{\hat{c}(t),\hat{c}^{\dagger}(0)\}\,|\Psi_0\rangle\,\theta(t), \qquad (3)$$

where $\{\,,\,\}$ denotes the anticommutator, $\Psi_0$ the electronic ground state, and the time-dependence of the operators is accounted for in the Heisenberg picture. As shown by Langreth (1970), the Green's function associated to the Hamiltonian in Eq. (1) can be calculated exactly. In fact, by applying a unitary transformation, Eq. (1) is recast in the form of a shifted harmonic oscillator Hamiltonian for which eigenvalues and eigenvectors are known (Mahan 2000). The spectral function for a localized electron can thus be expressed as (Langreth 1970):

$$A(\omega) = \sum_{n=0}^{\infty} \frac{e^{-a}a^n}{n!}\delta(\hbar\omega - \varepsilon - a\hbar\omega_{\mathrm{b}} + n\hbar\omega_{\mathrm{b}}), \qquad (4)$$

where $a = \sum_{\mathbf{q}} g_{\mathbf{q}}^2/(\hbar\omega_{\mathrm{b}})^2$ and for simplicity the energy of the boson mode has been replaced by its average value $\hbar\omega_{\mathrm{b}}$.

In the small coupling limit, that is for $g_{\mathbf{q}} \to 0$, the spectral function reduces to the case of a noninteracting electron $A(\omega) = \delta(\hbar\omega - \varepsilon)$, and the Dirac delta function is peaked at the quasiparticle energy. For finite coupling strengths, the structure of the spectral function in Eq. (4) reveals that the effect of the interaction between electrons and bosons on the spectral properties of the system is twofold. First, the quasiparticle energy of the localized electron is shifted by $a\hbar\omega_{\mathrm{b}}$. This process is analogous, for example, to the well-known band-gap renormalization of semiconductors and insulators due to the electron-phonon interaction (Allen and Heine 1976; Giustino 2017), and it results from the *dressing* of the bare quasiparticle via the interaction with the boson modes. Second, the spectral function exhibits a series of additional features at lower energies which are separated from the quasiparticle peak by multiples of the boson energy $\hbar\omega_{\mathrm{b}}$. These spectral features arise from the simultaneous excitation of the localized electron and of one or more bosons with energy $\hbar\omega_{\mathrm{b}}$.
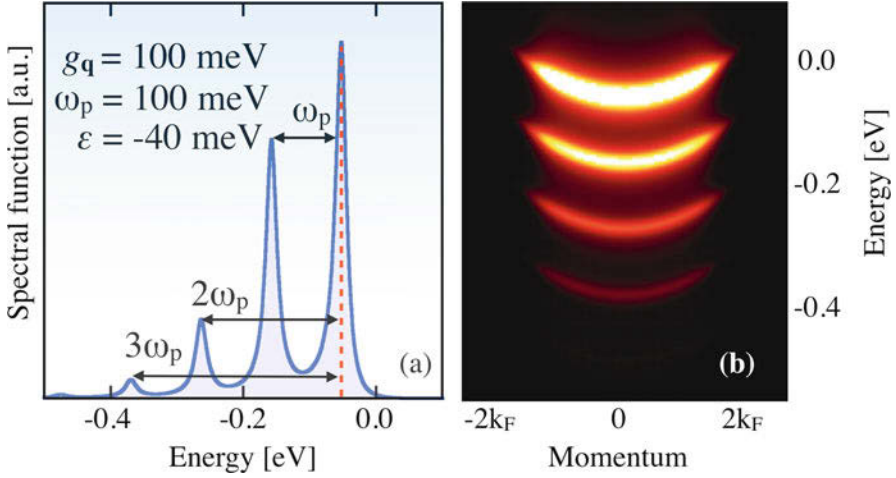
**Fig. 1** (**a**) Spectral function of the localized electron model evaluated using Eq. (4) for a nondispersive electron with binding energy $\varepsilon$ coupled to a boson with frequency $\omega_b$ with a coupling strength $g_q$. (**b**) Spectral intensity map for an electron with parabolic band dispersion $\varepsilon_k = \varepsilon + \hbar^2 k^2 / 2m$. The Dirac $\delta$ functions in Eq. (4) have been replaced by Lorentzian functions with a 20 meV broadening

In Fig. 1a we show the spectral function obtained from Eq. (4) considering $\varepsilon = -40\,\text{meV}$, $\hbar\omega_b = 100\,\text{meV}$, $g_q = 100\,\text{meV}$. A picture in closer agreement with angle-resolved photoemission spectroscopy is obtained when considering the case of dispersive electronic states: Fig. 1b illustrates the spectral intensity map for a parabolic band, obtained by replacing the electronic energy with $\varepsilon_k = \varepsilon + \hbar^2 k^2 / 2m$ in Eq. (4). This simple generalization of the localized electron model illustrates that, in presence of nontrivial energy-wavevector dispersion relations, the energy of the satellite features induced by electron-boson coupling also acquires a dependence on the crystal momentum that follows closely the dispersion of the ordinary quasiparticle states. This phenomenon translates into the formation of plasmonic polaron bands due to electron-plasmon coupling (Caruso et al. 2015; Caruso and Giustino 2015; Lischner et al. 2015; Gumhalter et al. 2016; Caruso et al. 2018) and polaron satellites due to electron-phonon coupling (Moser et al. 2013; Wang et al. 2016; Verdi et al. 2017) in the ARPES spectra of semiconductors and $n$-doped oxides, respectively.

Overall, the solution of the localized electron model reveals that the spectral function of a system of interacting electrons and bosons, whereby the interaction is described by the last term of Eq. (1), may exhibit a series of satellite structures, with binding energy blueshifted with respect to the main quasiparticle peak by multiples of the boson energy. Despite the simplicity of the model, this result provides a first indication that the coupling to plasmons and phonons in real systems, in which the coupling Hamiltonian assumes a similar form, may also induce the formation of satellite features for sufficiently strong coupling.

# 3      First-Principles Description of Satellites in Photoemission

Despite the different nature of plasmon and phonon collective excitations in solids, the many-body theory of electron-boson interaction represents the common playground to describe their coupling to electronic states and to investigate the spectral fingerprints resulting from this interaction. The Hedin-Baym equations (Giustino 2017) based on many-body perturbation theory (MBPT) provide a formally exact framework to investigate the coupling to plasmons and phonons and are the starting point for the theoretical description of satellites in photoemission spectra. The electron self-energy for the coupled electron-phonon system in the Migdal approximation, that is neglecting vertex corrections, is given by Hedin and Lundqvist (1970) and Giustino (2017):

$$\Sigma(\mathbf{k}, \omega) = i \int \frac{d\mathbf{q}}{\tilde{\Omega}} \frac{d\omega}{2\pi} G(\mathbf{k} + \mathbf{q}, \omega + \omega')[W_e(\mathbf{q}, \omega') + W_{ph}(\mathbf{q}, \omega')] \tag{5}$$

where $\tilde{\Omega}$ is the reciprocal-space volume, $G$ is the single-particle Green's function, and $W_e$ ($W_{ph}$) is the screened Coulomb interaction due to the electron-electron (electron-phonon) interaction. It can be shown that the electron-phonon part may be expressed as $W_{ph} = W_e D W_e$ (in symbolic notation), where $D$ is the density-density correlation function for the nuclear fluctuations. Eq. (5) neglects the so-called Debye-Waller contribution to the self-energy; however this contribution is frequency-independent and therefore does not give rise to additional structures in the electron spectral function (Giustino 2017).

If the nuclei are treated in the clamped-ion approximation, that is $W_{ph}(\mathbf{q}, \omega) = 0$, the ordinary $GW$ approximation is recovered. By expanding the Bloch wave functions in a basis set of plane waves, $\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_\mathbf{G} c_{n\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G}) \cdot \mathbf{r}}$, the $GW$ self-energy $\Sigma^{GW}$ can be expressed in a form more suitable for first-principles calculations of crystalline solids (Aulbur et al. 2000):

$$\Sigma_{n\mathbf{k}}^{GW}(\omega) = \frac{i\hbar}{2\pi} \sum_{m\mathbf{G}\mathbf{G}'} \int \frac{d\mathbf{q}}{\Omega_{BZ}} M_\mathbf{G}^{mn}(\mathbf{k}, \mathbf{q})^* M_{\mathbf{G}'}^{mn}(\mathbf{k}, \mathbf{q}) \int d\omega' \frac{v_\mathbf{G}(\mathbf{q}) \varepsilon_{\mathbf{G},\mathbf{G}'}^{-1}(\mathbf{q}, \omega)}{\hbar\omega + \hbar\omega' - \tilde{\varepsilon}_{m\mathbf{k}+\mathbf{q}}},$$
$$\tag{6}$$

where $M_\mathbf{G}^{mn}(\mathbf{k}, \mathbf{q}) = \langle \psi_{m\mathbf{k}+\mathbf{q}} | e^{i(\mathbf{q}+\mathbf{G}) \cdot \mathbf{r}} | \psi_{n\mathbf{k}} \rangle$ are the optical matrix elements, $\Omega_{BZ}$ is the volume of the Brillouin zone, and $v_\mathbf{G}(\mathbf{q}) = e^2/\varepsilon_0 |\mathbf{q} + \mathbf{G}|^2$ ($\varepsilon 0$ is the vacuum permittivity). We defined $\tilde{\varepsilon}_{m\mathbf{k}+\mathbf{q}} = \varepsilon_{m\mathbf{k}+\mathbf{q}} - i\eta \, \text{sgn}(\varepsilon_{m\mathbf{k}+\mathbf{q}})$, with $\eta$ a positive infinitesimal and $\varepsilon_{m\mathbf{k}+\mathbf{q}}$ the Bloch electron energy relative to the chemical potential $\mu$. The dielectric matrix $\varepsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega)$ is related to the screened Coulomb interaction via $W_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = v_\mathbf{G}(\mathbf{q}) \varepsilon_{\mathbf{G},\mathbf{G}'}^{-1}(\mathbf{q}, \omega)$. In $GW$ calculations, the dielectric function is typically expressed as $\varepsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = \delta_{\mathbf{G},\mathbf{G}'} - v_\mathbf{G}(\mathbf{q}) \chi_{\mathbf{G},\mathbf{G}'}^0(\mathbf{q}, \omega)$, where $\chi_{\mathbf{G},\mathbf{G}'}^0$ is the independent-particle polarizability (see, e.g., Aulbur et al. 2000).

The second term in Eq. (5) represents the electron-phonon self-energy $\Sigma^{\text{ep}}$ in the Migdal approximation. Its expression in the basis of single-particle Bloch wave functions reads:

$$\Sigma_{n\mathbf{k}}^{\text{ep}}(\omega) = \sum_{m\nu} \int \frac{d\mathbf{q}}{\Omega_{\text{BZ}}} |g_{mn\nu}(\mathbf{k}, \mathbf{q})|^2 \times$$

$$\left[ \frac{n_{\mathbf{q}\nu} + f_{m\mathbf{k}+\mathbf{q}}}{\hbar\omega - \varepsilon_{m\mathbf{k}+\mathbf{q}} + \hbar\omega_{\mathbf{q}\nu} + i\eta} + \frac{n_{\mathbf{q}\nu} + 1 - f_{m\mathbf{k}+\mathbf{q}}}{\hbar\omega - \varepsilon_{m\mathbf{k}+\mathbf{q}} - \hbar\omega_{\mathbf{q}\nu} + i\eta} \right], \qquad (7)$$

where $n_{\mathbf{q}\nu}$ and $f_{m\mathbf{k}+\mathbf{q}}$ are the Bose-Einstein and Fermi-Dirac distributions, respectively. Equation (7) is derived after transforming the frequency integration in Eq. (5) into a Matsubara summation to extend the formalism to finite temperatures, and performing the integration analytically by using the expressions for the unperturbed electron and phonon Green's functions. The self-energy is then analytically continued to the real frequency axis (Mahan 2000), and only the diagonal terms are retained, as in Eq. (6). The electron-phonon matrix element $g$ is defined as:

$$g_{mn\nu}(\mathbf{k}, \mathbf{q}) = \langle \psi_{m\mathbf{k}+\mathbf{q}} | \Delta_{\mathbf{q}\nu} V_{\text{KS}} | \psi_{n\mathbf{k}} \rangle. \qquad (8)$$

and it contains the variation of the self-consistent Kohn-Sham (KS) (Kohn and Sham 1965) potential $V_{\text{KS}}$ with respect to a phonon perturbation. The umklapp processes are included by letting $\mathbf{k} + \mathbf{q}$ fall outside the first Brillouin zone and folding it back with a reciprocal lattice vector $\mathbf{G}$. The definition in Eq. (8) corresponds to taking the bare Coulomb potential between the electrons and the nuclei screened by the electronic dielectric function $\varepsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega)$. In principle the matrix element should be frequency dependent; however in ab initio calculations, it is taken to be static, following the adiabatic approximation of standard density-functional theory (DFT). In Sect. 5 we will discuss how going beyond this approximation is needed when describing polarons in ARPES spectra. In practical calculations Eq. (8) is evaluated using density-functional perturbation theory (DFPT) by determining the linear variation of the self-consistent Kohn-Sham potential. A rigorous discussion of the calculation of the DFPT screening as compared to the many-body random-phase approximation (RPA) can be found, for example, in Marini et al. (2015).

## 3.1   The Electron Spectral Function

The calculation of the self-energies defined by Eqs. (6) and (7) constitutes the first step toward the description of satellites from first principles. Details regarding the numerical evaluation of these expressions have been thoroughly reported, for instance, in Marini et al. (2009) and Poncé et al. (2016) and will not be discussed here. Once the electron self-energy $\Sigma_{n\mathbf{k}}(\omega)$ is known, the spectral function is obtained by combining Eq. (2) with the Dyson's equation $G_{n\mathbf{k}} = [\hbar\omega - \varepsilon_{n\mathbf{k}} - \Sigma_{n\mathbf{k}}(\omega)]^{-1}$, which yields:

$$A(\mathbf{k}, \omega) = -\frac{1}{\pi} \sum_n \frac{\operatorname{Im} \Sigma_{n\mathbf{k}}(\omega)}{[\hbar\omega - \varepsilon_{n\mathbf{k}} - \operatorname{Re} \Sigma_{n\mathbf{k}}(\omega)]^2 + [\operatorname{Im} \Sigma_{n\mathbf{k}}(\omega)]^2} \tag{9}$$

The spectral function exhibits sharp peaks whenever the first term in the denominator of Eq. (9) $[\hbar\omega - \varepsilon_{n\mathbf{k}} - \operatorname{Re} \Sigma_{n\mathbf{k}}(\omega)]$ vanishes or has a minimum. In particular, quasiparticle peaks in the spectral function arise at the energies $\hbar\omega = \varepsilon_{n\mathbf{k}} + Z_{n\mathbf{k}} \operatorname{Re} \Sigma_{n\mathbf{k}}(\varepsilon_{n\mathbf{k}})$, where $Z_{n\mathbf{k}} = [1 - \partial \operatorname{Re} \Sigma_{n\mathbf{k}}(\omega)/\partial\omega|_{\omega=\varepsilon_{n\mathbf{k}}}]^{-1}$ is the quasiparticle weight. If the Bloch single-particle energies $\varepsilon_{n\mathbf{k}}$ are obtained from a DFT calculation, Eq. (9) should be modified to avoid double counting of the exchange-correlation (Aulbur et al. 2000).

A more suitable framework for the evaluation of satellites in photoemission is provided by the cumulant expansion approach. The cumulant expansion is an alternative formulation of the (retarded) single-particle Green's function which is in principle exact. The Green's function is expressed in the form (Gumhalter et al. 2016; Kas et al. 2014):

$$G_{n\mathbf{k}}(t) = i\theta(t)\exp[-i(\varepsilon_{n\mathbf{k}} - i\eta)t/\hbar + C_{n\mathbf{k}}(t)], \tag{10}$$

where we introduced the cumulant function $C_{n\mathbf{k}}(t)$ which is defined by:

$$C_{n\mathbf{k}}(t) = -\frac{1}{\hbar\pi} \int d\omega \operatorname{Im} \Sigma_{n\mathbf{k}}(\varepsilon_{n\mathbf{k}}/\hbar - \omega) \frac{1 - e^{i\omega t} + i\omega t}{\omega^2} \tag{11}$$

In practice, the spectral function obtained from Eqs. (2), (10), and (11) can be recast into a form that is more suitable for numerical calculations (Aryasetiawan et al. 1996; Aryasetiawan and Gunnarson 1998; Verdi et al. 2017):

$$A(\mathbf{k}, \omega) = \sum_n \left[ 1 + A_{n\mathbf{k}}^{S1}(\omega) * + \frac{1}{2} A_{n\mathbf{k}}^{S1}(\omega) * A_{n\mathbf{k}}^{S1}(\omega) * + \cdots \right] A_{n\mathbf{k}}^{QP}(\omega). \tag{12}$$

Here we introduced the following quantities:

$$A^{QP}(\omega) = \frac{e^{\operatorname{Re}\Sigma'(\varepsilon/\hbar)}}{\pi} \frac{|\operatorname{Im}\Sigma(\varepsilon/\hbar)|}{[\hbar\omega - \varepsilon - \operatorname{Re}\Sigma(\varepsilon/\hbar)]^2 + [\operatorname{Im}\Sigma(\varepsilon/\hbar)]^2},$$

$$A^{S1}(\omega) = -\frac{1}{\pi} \frac{\operatorname{Im}\Sigma(\varepsilon/\hbar + \omega) - \operatorname{Im}\Sigma(\varepsilon/\hbar) - \hbar\omega\operatorname{Im}\Sigma'(\varepsilon/\hbar)}{(\hbar\omega)^2},$$

where we omitted the dependence on $n$ and $\mathbf{k}$, and the prime symbol denotes the first derivative. In the limit of a localized electron interacting with a plasmon bath, one may show that Eq. (12) reduces to the exact solution of the localized electron model given by Eq. (4). The application of this formalism to core and valence excitations of crystalline solids, on the other hand, involves several approximations such as neglecting *recoil* effects, that is, the correlations between successive boson

emission and reabsorption events. A detailed discussion of the range of validity of the cumulant expansion has been reported, for instance, in Hedin (1980), Gumhalter et al. (2016), Kas et al. (2014), and Zhou et al. (2015).

As discussed by Holm and Aryasetiawan (1997), the cumulant expansion has the advantage of introducing additional crossing and non-crossing Feynman diagrams that are neglected in the standard $GW$ and Migdal approximation for the self-energy, and it results in an improved description of the electron-plasmon and electron-phonon interactions. The ab initio cumulant expansion approach is based on the evaluation of Eq. (12) employing either the $GW$ or Migdal self-energy. As discussed in Sect. 6, this formalism also lends itself to describe the formation of hybrid modes resulting from the combined effects of plasmons and phonons.

The self-consistent solution of the Dyson's equation could in principle provide an alternative route to include additional diagrams beyond the $GW$/Migdal approximation in the Green's function. However, while self-consistent $GW$ has been shown to systematically improve the description of the quasiparticle energies of molecules and solids (Caruso et al. 2013, 2016; Kutepov et al. 2012), the study of satellites has revealed that self-consistency leads to an unphysical renormalization of the satellite intensity which, ultimately, is expected to deteriorate the agreement with experiment (Holm and von Barth 1998). Additional first-principles investigations would be needed to further explore this aspect.

## 3.2   Plasmon Satellites

The concept of plasmons, collective fluctuations of the electron density, can be introduced based on a simple model of carrier dynamics for a homogeneous system (that is, a system characterized by a homogeneous electron density and a positively charged ionic background) in which the quantum-mechanical character of the electrons is ignored. If an external perturbation as, for example, a homogeneous electric field is present, a displacement $\mathbf{x}$ of the electron density with respect to the positively charged ionic background is induced. The displaced electron density then generates an induced polarization $\mathbf{P} = -ne\mathbf{x}$, where $n$ is the average electron density and $e$ the electron charge and an electric field $\mathbf{E} = -\mathbf{P}/\varepsilon_0$. Using Newton's law $m\ddot{\mathbf{x}} = -e\mathbf{E}$ with $m$ the electron mass, the classical equation of motion for the density displacement vector $\mathbf{x}$ may be rewritten as $\ddot{\mathbf{x}} + ne^2\mathbf{x}/(\varepsilon_0 m) = 0$. This model illustrates that the classical collective dynamics of electrons in solids can be approximately described by a harmonic oscillator with a characteristic frequency $\omega_P = \sqrt{ne^2/(\varepsilon_0 m)}$, the plasma frequency, which is independent of the perturbation and is determined exclusively by the intrinsic properties of the solid.

More generally, plasmons in solids may be excited at momenta $\mathbf{q}$ and frequencies $\omega_P$ which correspond to vanishing real part of the macroscopic dielectric function $\varepsilon_M$ and sufficiently small imaginary part, that is:

$$\varepsilon_M(\mathbf{q}, \omega_P) = i\eta, \tag{13}$$
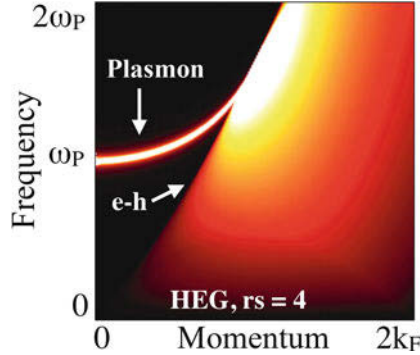
**Fig. 2** The loss function of the HEG for a Wigner-Seitz radius $r_s = 4$. The plasmonic structures in the loss function follow a characteristic parabolic dispersion which, for momenta smaller than a critical momentum $q_c$, is well separated from the continuum of electron-hole excitations (e-h). For $q > q_c$, the plasmons are damped by the interaction with electron-hole pairs (Landau damping)

The macroscopic dielectric function $\varepsilon_M$ is related to the microscopic dielectric function $\varepsilon$ via $\varepsilon_M^{-1}(\mathbf{q}, \omega) = [\varepsilon_{\mathbf{G},\mathbf{G'}}(\mathbf{q}, \omega)]_{\mathbf{G}=\mathbf{G'}=0}^{-1}$, with $\mathbf{G}$ and $\mathbf{G'}$ reciprocal lattice vectors and $\mathbf{q}$ in the first Brillouin zone. Whenever the condition expressed by Eq. (13) is satisfied, the system may support collective charge fluctuations even in the absence of an external driving field. In practice, the possibility of exciting plasmons is reflected by the emergence of sharp peaks in the loss function $L(\mathbf{q}, \omega) = \text{Im}\,[\varepsilon_M^{-1}(\mathbf{q}, \omega)]$ at the momenta and frequencies at which the macroscopic dielectric function $\varepsilon_M$ vanishes. The plasmon peaks in the loss function exhibit well-defined energy-momentum dispersion relations. These structures are exemplified in Fig. 2 for the loss function of the homogeneous electron gas (HEG). If one neglects local-field effects and thus assumes that the macroscopic and microscopic dielectric functions coincide, the plasmon energy is obtained by seeking the frequencies that satisfy the condition $v(\mathbf{q})^{-1} = \text{Re}\,\chi_0(\mathbf{q}, \omega)$. For the HEG in the long-wavelength limit ($\mathbf{q} \to 0$), this condition yields again the result $\omega_P = \sqrt{e^2 n/(\varepsilon_0 m)}$.

The inspection of Eqs. (13) and (6) reveals that when the condition for the excitation of plasmons is satisfied, the screened Coulomb interaction $W_{\mathbf{G},\mathbf{G'}}(\mathbf{q}, \omega) = v_{\mathbf{G}}(\mathbf{q})[\varepsilon_{\mathbf{G},\mathbf{G'}}(\mathbf{q}, \omega)]^{-1}$ exhibits a pole at the plasmon energy. Correspondingly, one expects the $GW$ self-energy to encode information regarding electron-plasmon interaction.

To examine in more detail the inclusion of electron-plasmon coupling effects in the $GW$ self-energy, we discuss below its connection with the electron-boson coupling model introduced in Sect. 2. Using the condition given in Eq. (13) in combination with Eq. (6), the plasmonic contribution to the screened Coulomb interaction $W$ can be disentangled from the other electronic contributions to the screening, such as electron-hole pairs. This idea, initially introduced for the homogeneous electron gas (Lundqvist 1967) and subsequently generalized to semiconductors (Caruso and Giustino 2016), allows one to define a self-energy

which stems exclusively from the coupling between electrons and plasmons. The resulting electron-plasmon self-energy can be recast into the following form (Caruso and Giustino 2016):

$$
\Sigma_{n\mathbf{k}}^{\mathrm{eP}} = \int \frac{d\mathbf{q}}{\Omega_{\mathrm{BZ}}} \sum_m |g_{mn}^{\mathrm{eP}}(\mathbf{k}, \mathbf{q})|^2
$$

$$
\times \left[ \frac{n_\mathbf{q} + f_{m\mathbf{k}+\mathbf{q}}}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}+\mathbf{q}} + \hbar\omega_\mathrm{P}(\mathbf{q}) + i\eta} + \frac{n_\mathbf{q} + 1 - f_{m\mathbf{k}+\mathbf{q}}}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}+\mathbf{q}} - \hbar\omega_\mathrm{P}(\mathbf{q}) + i\eta} \right],
\tag{14}
$$

where the coefficients $g_{mn}^{\mathrm{eP}}(\mathbf{k}, \mathbf{q})$ are the electron-plasmon scattering matrix elements between the initial state $\psi_{n\mathbf{k}}$ and the final state $\psi_{m\mathbf{k}+\mathbf{q}}$ and are given by:

$$
g_{mn}^{\mathrm{eP}}(\mathbf{k}, \mathbf{q}) = \left[ \frac{\varepsilon_0 \Omega}{e^2 \hbar} \frac{\partial \varepsilon(\mathbf{q}, \omega)}{\partial \omega} \right]_{\omega_\mathrm{P}(\mathbf{q})}^{-\frac{1}{2}} \frac{1}{|\mathbf{q}|} \langle \psi_{m\mathbf{k}+\mathbf{q}} | e^{i\mathbf{q}\cdot\mathbf{r}} | \psi_{n\mathbf{k}} \rangle,
\tag{15}
$$

with $\Omega$ being the volume of the unit cell. Equation (14) has the form of an electron-boson coupling self-energy in the Migdal approximation (see Eq. (7)), which may alternatively be derived from an electron-boson coupling Hamiltonian of the form:

$$
\hat{H}^{\mathrm{eP}} = \sum_{nm} \sum_{\mathbf{k}, \mathbf{q}} g_{nm}^{\mathrm{eP}}(\mathbf{k}, \mathbf{q}) \hat{c}_{m\mathbf{k}+\mathbf{q}}^\dagger \hat{c}_{n\mathbf{k}} (\hat{b}_\mathbf{q} + \hat{b}_{-\mathbf{q}}^\dagger).
\tag{16}
$$

Here $\hat{b}_{-\mathbf{q}}^\dagger$ ($\hat{b}_\mathbf{q}$) and $\hat{c}_{m\mathbf{k}+\mathbf{q}}^\dagger$ ($\hat{c}_{n\mathbf{k}}$) are the boson and fermion creation (destruction) operators, respectively. The localized electron model of Eq. (1) is recovered from Eq. (16) by (i) replacing the Bloch energies $\varepsilon_{n\mathbf{k}}$ with a single nondispersive energy and (ii) neglecting the $\mathbf{k}$-dependence of the electron-boson coupling matrix elements. This result indicates that the $GW$ self-energy accounts for the coupling between electrons and plasmons. However, at variance with the localized electron model which could be solved exactly, here the electron-plasmon interaction is treated only at first-order in the interaction strength, which corresponds to the Migdal approximation in the ordinary electron-boson coupling theory.

The inclusion of electron-plasmon coupling in the $GW$ theory is reflected by the emergence of plasmon satellites in the spectral function, which are analogous to the satellite features discussed in Sect. 2. In fact, in the presence of plasmons, the frequency dependence of the self-energy typically presents a pole, which may produce additional satellite structures in the spectral function signaling the coupling to plasmons.

Two clear shortcomings emerge when evaluating the spectral function within the $GW$ approximation and limit its predictive power for the description of satellites in PES: (i) the energy difference between the satellite and the quasiparticle peak is typically overestimated by a factor of 1.5 with respect to photoemission experiments, and (ii) the $GW$ approximation may erroneously predict the formation of
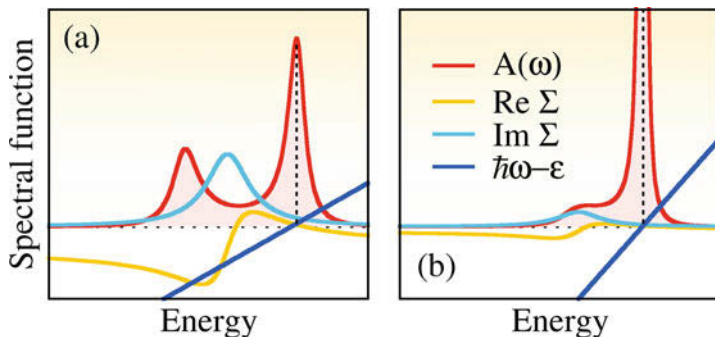
**Fig. 3** Spectral function for a model self-energy in the strong (**a**) and weak (**b**) coupling regime. Quasiparticle peaks are marked by vertical dashed lines and correspond to the intersection between $\mathrm{Re}\,\Sigma(\omega)$ and $\hbar\omega - \varepsilon$

spurious *plasmaron* peaks, which stem from additional solutions of the quasiparticle equation and often result in an overestimation of satellite intensities. The concept of plasmaron was initially introduced by Lundqvist (1967) as a new quasiparticle state emerging from the strong coupling between electrons and plasmons. Later studies, however, revealed that plasmaron peaks are an artifact of the $GW$ approximation and, in fact, they disappear when one resorts to a more accurate level of theory (Langreth 1970). These issues can be illustrated by using a simplified model for the $GW$ self-energy: $\Sigma(\omega) = \alpha(\hbar\omega - \varepsilon + \omega_P + i\eta)^{-1}$. This expression is derived from Eq. (6) by (i) assuming nondispersive electron energies, (ii) replacing the oscillator strengths by $\delta$ functions, (iii) using a plasmon-pole model for the dielectric function in the form $\varepsilon^{-1}(\omega) = 1 + \tilde{\omega}/(\omega^2 - \omega_P^2 + i\eta)$, and (iv) carrying out the frequency integration analytically. As shown in Fig. 3, the self-energy exhibits a pole at frequencies around $\omega = \hbar^{-1}(\varepsilon - \hbar\omega_P)$, which may lead to additional unphysical solutions of the quasiparticle equation when $\hbar\omega - \varepsilon = Re\,\Sigma(\omega)$ as shown in panel (a) or to a minimum in $\hbar\omega - \varepsilon - Re\,\Sigma(\omega)$ resulting in a weak satellite, as shown in panel (b). In both cases the spectral function is characterized by the emergence of satellites; however their binding energy is blue-shifted with respect to the energy $\varepsilon - \hbar\omega_P$ at which satellites are typically observed in PES experiment. As we will discuss in Sect. 4, the combination of the $GW$ approximation with the cumulant expansion approach ($GW$+C) allows to successfully address these issues and recover an energy separation between satellite and quasiparticle peaks that agrees well with PES measurements for a broad class of materials.

## 3.3   Polaron Satellites

Similarly to the case of the interaction with plasmons, the coupling between electrons and phonons may give rise to satellite structures in the spectral function of semiconductors and insulators, which are the signature of the dressing of the electronic quasiparticles as *polarons*. The formation of polarons is typically linked

to the polarization of the lattice induced by longitudinal optical (LO) phonons. In other words, in polar semiconductors and insulators, the fluctuations of the ionic positions corresponding to LO phonons at long wavelength generate macroscopic electric fields which can couple strongly to electrons and holes. This long-range interaction is known as Fröhlich coupling. The Fröhlich model strictly describes the interaction of a conduction electron in a parabolic band with LO phonons of constant energy $\hbar\omega_{LO}$, in an isotropic and uniform medium (Fröhlich 1954). Under these assumptions, the electron-phonon matrix element does not depend on the band index and electron momentum, and it takes the form:

$$g_F(\mathbf{q}) = \frac{i}{|\mathbf{q}|} \left[ \frac{e^2}{4\pi\varepsilon_0} \frac{4\pi}{\Omega} \frac{\hbar\omega_{LO}}{2} \left( \frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_s} \right) \right]^{1/2}, \tag{17}$$

where $\varepsilon_s$ is the total static permittivity (or dielectric constant) and $\varepsilon_\infty$ is the optical dielectric constant, that is, $\varepsilon_M(\mathbf{q} = 0, \omega = 0)$. The matrix element in Eq. (17) is often expressed in terms of a dimensionless parameter $\alpha$ which is referred to as the *Fröhlich coupling constant*:

$$\alpha = \frac{e^2}{\hbar} \left( \frac{m_b}{2\hbar\omega_{LO}} \right)^{1/2} \left( \frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0} \right), \tag{18}$$

with $m_b$ the band effective mass of the conduction electron. The Fröhlich Hamiltonian has the form in Eq. (16) after substituting $g_{mn}^{eP}$ with $g_F\delta_{mn}$, and it is thus historically representative of the general problem of a fermionic particle interacting with a boson field. Depending on the value of $\alpha$, i.e., on the strength of the coupling, the Fröhlich self-energy produces a spectral function that usually exhibits satellite replica of the main quasiparticle band.

First-principles calculations of electron-phonon self-energies and ARPES spectra to capture polaronic effects are limited by the almost prohibitive computational cost of sampling the singular behavior of the matrix elements for small phonon wavevectors. A procedure that enables accurate calculations of the electron-phonon coupling in the presence of Fröhlich interaction at a reduced computational cost has been reported by Verdi and Giustino (2015) and Sjakste et al. (2015). This is achieved via the separation of the long-range, singular part of the electron-phonon matrix element and of the short-range part. The long-range part $g^{\mathcal{L}}$ constitutes the generalization of the Fröhlich matrix element to multiple, anisotropic electronic bands and phonon modes and reads:

$$g_{mn\nu}^{\mathcal{L}}(\mathbf{k}, \mathbf{q}) = i \frac{4\pi e^2}{\Omega} \sum_\kappa \left( \frac{\hbar}{2M_\kappa\omega_{\mathbf{q}\nu}} \right)^{1/2}$$
$$\times \sum_{\mathbf{G}, \mathbf{q}+\mathbf{G}\neq 0} \frac{(\mathbf{q}+\mathbf{G}) \cdot \mathbf{Z}_\kappa^* \cdot \mathbf{e}_{\kappa\nu}(\mathbf{q})}{(\mathbf{q}+\mathbf{G}) \cdot \boldsymbol{\varepsilon}_\infty \cdot (\mathbf{q}+\mathbf{G})} \langle \psi_{m\mathbf{k}+\mathbf{q}} | e^{i(\mathbf{q}+\mathbf{G}) \cdot (\mathbf{r}-\tau_\kappa^0)} | \psi_{n\mathbf{k}} \rangle, \tag{19}$$

where $\mathbf{Z}_\kappa^*$ is the Born effective charge tensor of atom $\kappa$ in the unit cell, $M_\kappa$ the atomic mass, and $\mathbf{e}_{\kappa\nu}(\mathbf{q})$ a phonon eigenvector. If combined with the Wannier-Fourier interpolation technique of Giustino et al. (2007), Eq. (19) enables accurate calculations of polaron satellites.

As in the case of the $GW$ method, the calculation of the spectral function including electron-phonon coupling in the Migdal approximation suffers from two main shortcomings. First, it produces only a single polaronic satellite rather than a Lang-Firsov series as shown by the model of Eqs. (1) and (4) and as measured in experiments. Second, its energy separation from the main quasiparticle peak is larger than the characteristic LO phonon energy. As we will illustrate in Sect. 4 and 5, the cumulant expansion method can successfully be employed to improve the description of satellites. While this method has been mostly used in combination with the $GW$ approximation to study plasmon satellites, it can also naturally be applied in the context of polaronic systems, since the theory stems from the exact solution of an electron-boson coupling Hamiltonian of the Fröhlich type (Langreth 1970; Story et al. 2014). The formalism corresponds to the one presented in Sect. 3.1, with the Migdal electron-phonon self-energy used as a seed.

## 4    Plasmon Satellites in Metals and Semiconductors

First-principles calculations of plasmon satellites based on the $GW$+C approach have first been performed by Aryasetiawan et al. (1996) for metallic sodium and aluminum. The integrated photoemission spectroscopy experiment on sodium by Steiner et al. (1979) revealed, besides a quasiparticle peak centered at a binding energy of 1 eV which corresponds to the excitation of photo-holes in the valence band, two broader and less intense satellite peaks blue-shifted with respect to the quasiparticle peak by 6 and 12 eV respectively. These energies are compatible with multiples of the plasma energy of sodium $\hbar\omega_\mathrm{P} \simeq 5.9$ eV, suggesting that the satellites arise from the excitation of one and two plasmons. At variance with the $GW$ results, which overestimate the energy and intensity of the satellites, the spectral function of Na obtained from the $GW$+C approach and shown in Fig. 4a improves significantly the agreement with the experiment. Additionally, the $GW$+C approach lends itself to describe also processes in which more than one plasmon are excited and captures the emergence of a series of satellite peaks spaced by the plasmon energy. On the other hand, only one satellite is obtained within the $GW$ approximation, reflecting the fact that multi-plasmon processes are neglected.

Subsequently, photoemission satellites have been measured in the photoemission spectra of graphene, and the identification of these features has been supported by theoretical calculations of the self-energy and spectral function for linearly dispersive bands (Bostwick et al. 2010). Satellites in semiconductors have first been investigated from first principles by Guzzo et al. (2011) for the case of silicon. In analogy with metals, also photoemission measurements of semiconductors may exhibit a series of satellites (Fig. 4b) with an energy separation that is compatible with the plasma energy. In this case, however, plasmons are generally characterized
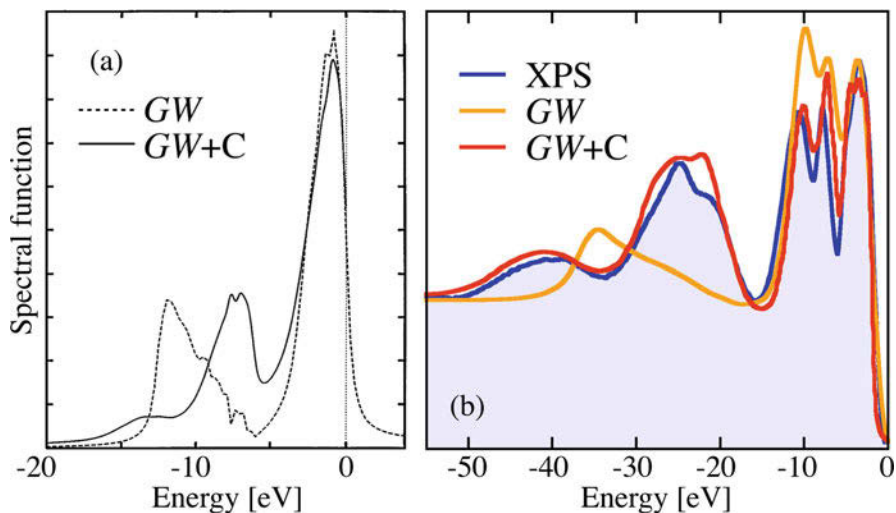
**Fig. 4**   (**a**) Cumulant expansion for Na. (**b**) Cumulant expansion for silicon (Reproduced with permission from Aryasetiawan et al. 1996 and Guzzo et al. 2011)

by a smaller oscillator strength, which is reflected by the lower intensity of the satellite peaks. For silicon, the $GW$ approximation yields a single satellite blue-shifted by $\sim$22 eV with respect to the quasiparticle peak, which is incompatible with the plasma energy $\hbar\omega_P = 16.6$ eV and with the experimental observations. On the other hand, when vertex corrections are included via the $GW$+C approach, the energy of the plasmon satellite peak is in good agreement with experiment. Some discrepancies between theory and experiment still remain, namely, (i) the intensity of the satellite peak is underestimated, (ii) the relative intensity between the different substructures of the quasiparticle peak differs from the experimental result, and (iii) experiments present a featureless background signal that increases with the electron binding energy and that is not captured by theory. The points (i)–(iii) are directly related to the interpretation of the spectral function as a photoelectron current, an approximation that is typically referred to as *sudden approximation* (Hüfner 2003; Damascelli et al. 2003). In practice, the sudden approximation assumes that all electrons are equally likely to interact with an incoming photon and that after photoexcitation the electrons do not interact further with the sample. These assumptions neglect the scattering cross-section effects due to the different orbital symmetries and the additional energy losses that photo-electrons may undergo after emission from the initial state. The issues mentioned in (i)–(iii) could be improved by adopting a picture of the photoemission process that goes beyond the sudden approximation, e.g., by explicitly accounting for extrinsic losses, background signal, and cross-section effects. In this way, a quantitatively accurate description of satellites in semiconductors may be achieved (Guzzo et al. 2012).

Inspection of the first satellite peak in Fig. 4b indicates that the plasmon-induced spectral features of silicon are characterized by a substructure – in this case a central peak and two shoulders observed in both theory and experiments – that resembles the density of states (DOS) of the ordinary quasiparticle bands. To understand the origin of these features, it is convenient to recall the concept of Van Hove singularities from the quantum theory of solids. The density of states $J$ for a set of Bloch electrons can be expressed as:

$$J(\omega) = \frac{1}{4\pi^3} \sum_n \int_{S(\hbar\omega)} dS_{\mathbf{k}} \frac{1}{|\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}|} \tag{20}$$

where the integral is performed over the isosurfaces in $\mathbf{k}$-space with energy $\hbar\omega$, denoted by $S(\hbar\omega)$. If for a given energy $\hbar\omega$, the isosurface $S(\hbar\omega)$ contains a crystal momentum for which the electron velocity vanishes ($v_{n\mathbf{k}} = \nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}}/\hbar = 0$), the divergence of the integrand in Eq. (20) leads to a sharp structure in $J(\omega)$, referred to as a Van Hove singularity. Peaks in the DOS may thus be attributed to regions of the Brillouin zone in which electronic bands are flat ($\nabla_{\mathbf{k}}\varepsilon_{n\mathbf{k}} \simeq 0$). These structures are clearly visible in PES experiments of silicon for binding energies between 0 and $-15$ eV (Fig. 5b) and in the DOS obtained from DFT calculations in the local density approximation (Fig. 5c). The structure of plasmon satellites measured in PES also exhibits a substructure of peaks and shoulders that resembles the Van Hove singularities, as it can be noted when comparing it with the DOS of the ordinary quasiparticle bands red-shifted by the plasmon energy (Fig. 5b–c). This suggests
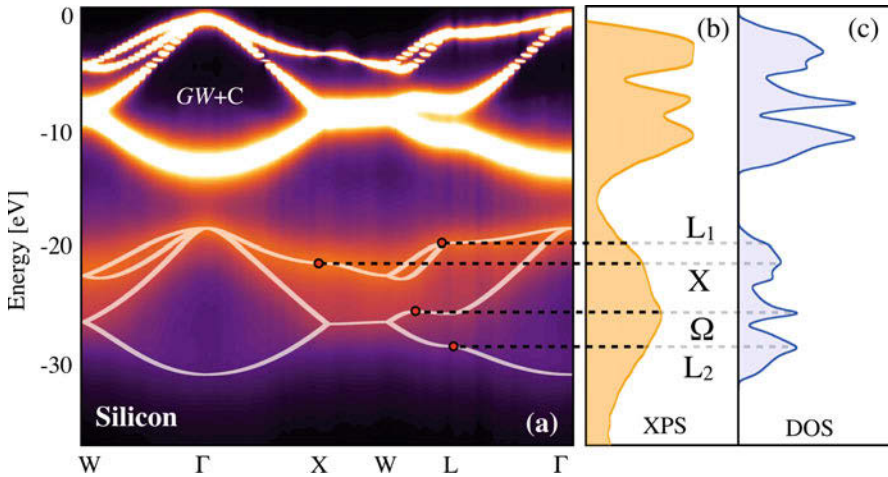


**Fig. 5** (**a**) Theoretical calculations of the plasmonic polaron band of silicon, based on the $GW$ plus cumulant expansion approach. (Adapted from Gumhalter et al. 2016). (**b**) Integrated X-ray photoemission spectrum (XPS) of silicon from Guzzo et al. (2011). (**c**) Density of states of silicon from a density-functional theory calculation alongside with a replica of the full DOS red-shifted by the plasmon energy $\hbar\omega_P = 16.6$ eV

that the plasmon satellites observed in integrated PES also stem from the average over the Brillouin zone of spectral features that are characterized by well-defined energy-momentum dispersion relations.

This hypothesis has been verified by first-principles calculations of the angle-resolved spectral function of silicon in the $GW$+C approach (Caruso et al. 2015), which revealed that electron-plasmon interaction leads to the emergence of *plasmonic polaron bands*, that is, plasmon-induced replica of the valence band structure of semiconductors red-shifted by the plasmon energies. These features are illustrated for silicon in Fig. 5a. As compared to the quasiparticle bands, plasmonic polaron bands are less intense due to the small oscillator strength of plasmon in semiconductors and broadened out by lifetime effects. The existence of plasmonic polaron bands has been corroborated by further theoretical and experimental investigations of the ARPES spectrum of silicon for binding energies up to 40 eV by Lischner et al. (2015). Overall, the dispersive character of plasmon-induced features in ARPES indicates that plasmon satellites in integrated PES may be interpreted as Van Hove singularities which arise from the flattening of the plasmonic polaron bands at specific regions in the Brillouin zone.

## 5    Polaron Satellites in Doped Semiconductors

Low-energy satellites have recently been observed by ARPES experiments in doped oxides. These systems constitute an ideal playground for the study of polaron physics. In particular, satellite replicas were measured for $n$-doped $TiO_2$ (Moser et al. 2013), $SrTiO_3$ (Chang et al. 2010), and monolayer FeSe on $SrTiO_3$ (Lee et al. 2014). Evidence of Fröhlich polarons was found also from the investigation of two-dimensional (2D) electronic states at the surfaces or interfaces of oxides, with the most studied case being the 2D electron gas (2DEG) formed at the surface of $SrTiO_3$ (King et al. 2014; Chen et al. 2015; Wang et al. 2016). Other notable examples are the 2DEG at the interface between $SrTiO_3$ and $LaAlO_3$ (Cancellieri et al. 2016) and on the surface of ZnO (Yukawa et al. 2016). The experiments also show a remarkable evolution of the carriers with doping concentration, from polarons to a Fermi liquid weakly coupled to phonons (Moser et al. 2013; Wang et al. 2016).

Calculations of the spectral function using model self-energies or the localized electron model of Eq. (4) have been performed for some of these systems, for example, in Moser et al. (2013), Lee et al. (2014), King et al. (2014), and Rademaker et al. (2016). Fully *ab initio* calculations showing satellite band replica were first reported for the insulating compounds MgO and LiF by Antonius et al. (2015), subsequently exploring also the effect of the cumulant expansion method (Nery et al. 2018). First-principles calculations of ARPES spectra in doped materials including polaronic effects were carried out by Verdi et al. (2017) for the prototypical case of anatase $TiO_2$ by using the methods presented in Sect. 3. Given that the crystals are doped, an important element that needs to be taken into account when performing *ab initio* calculations is the presence of additional charges in the conduction band. Since the systems of interest are degenerate and present well-

defined Fermi surfaces, doping can be treated in the rigid-band approximation, that is, by placing the Fermi level inside the conduction or valence band of the pristine system. Moreover, the added carriers provide an additional source of screening of the electron-phonon interactions. This effect is critical especially in the case of polar coupling, where the screening of the macroscopic electric field created by the LO phonons can change dramatically the strength of the Fröhlich interaction (Mahan 2000). This aspect can be understood by considering the simple Thomas-Fermi screening model, which describes the static response of a homogeneous electron gas at small wavevectors: $\varepsilon_{\mathrm{TF}}(\mathbf{q}) = 1 + q_{\mathrm{TF}}^2/|\mathbf{q}|^2$, with $q_{\mathrm{TF}} = \sqrt{2e^2 n/(\varepsilon_0 \varepsilon_\infty E_{\mathrm{F}})}$ ($E_{\mathrm{F}}$ is the Fermi energy). From the wavevector dependence of $\varepsilon_{\mathrm{TF}}(\mathbf{q})$, it follows immediately that the screened matrix element, $g_{\mathrm{F}}(\mathbf{q})/\varepsilon_{\mathrm{TF}}(\mathbf{q})$, no longer exhibits a singularity at long wavelength. This model is valid in the adiabatic limit where the doped carriers instantaneously follow the atomic motion. In a more accurate description, the timescale of the electronic response is dictated by the plasma frequency of the doped carriers, $\omega_{\mathrm{P}} = \sqrt{ne^2/(\varepsilon_0 \varepsilon_\infty m_{\mathrm{b}})}$ in the HEG model (Kittel 1976).

To capture the evolution of the electron-phonon coupling and of the polaronic features with doping, the electron-phonon matrix element needs to be screened by the dynamical dielectric function evaluated at the phonon energies, that is, $g_{mn\nu}^{\mathrm{NA}}(\mathbf{k}, \mathbf{q}) = g_{mn\nu}(\mathbf{k}, \mathbf{q})/\varepsilon_{\mathrm{RPA}}(\mathbf{q}, \omega_{\mathbf{q}\nu} + i/\tau_{n\mathbf{k}})$ (Mahan 2000; Verdi et al. 2017). The superscript NA indicates that retardation effects are taken into account by using this non-adiabatic matrix element, and $\hbar/\tau_{n\mathbf{k}}$ is the electron lifetime near the band edge, which can approximately be taken to be constant. In practical calculations the dynamical screening arising from the doped carriers can be computed analytically using the RPA dielectric function for a homogeneous electron gas with the same density $n$, which is known as the Lindhard function (Hedin 1965).

In Fig. 6 we show the ARPES spectra acquired for $n$-doped anatase $TiO_2$ by Moser et al. (2013), and we compare them with the first-principles calculations performed by Verdi et al. (2017). The spectra for the first two doping levels exhibit a satellite about 0.1 eV below the main parabolic band and a second very dim satellite at another 0.1 eV higher binding energy. Since the energy separation of the band replica is compatible with the high-energy $E_u$ LO phonon of anatase $TiO_2$, these satellites were attributed to polaronic effects. At the highest doping, on the other hand, the satellites disappear and are replaced by band structure kinks near a binding energy of 0.1 eV. All the spectral features and their evolution with doping are reproduced by the calculations, thus confirming the transition from a polaronic to a Fermi liquid picture and demonstrating the success of the first-principles methods used to investigate quasiparticle spectra. From the calculated ARPES spectra, the electron-phonon coupling strength $\lambda$ was extracted, by using the ratio between the Fermi velocities of the bare band and of the dressed band (Mahan 2000). The results are reported in Fig. 7, together with an analysis of the energy scales at play. The study showed that the crossover from polarons to a weakly coupled Fermi liquid and, correspondingly, from satellite replica to band structure kinks occurs when the plasma frequency of the carriers becomes of the order of the LO phonon frequency. In fact, in the polaronic regime, corresponding to $\omega_{\mathrm{P}} < \omega_{\mathrm{LO}}$, the carriers
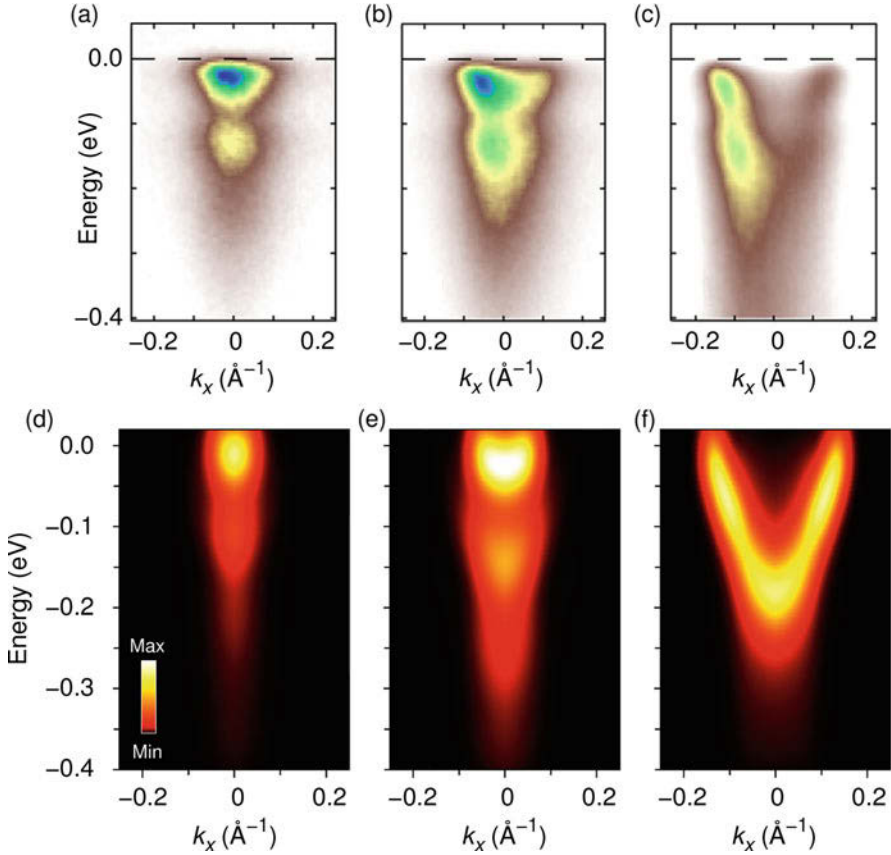
**Fig. 6** ARPES spectra of *n*-doped anatase $TiO_2$ on samples with doping concentrations $5 \times 10^{18}$ cm$^{-3}$ (**a**), $3 \times 10^{19}$ cm$^{-3}$ (**b**), and $3.5 \times 10^{20}$ cm$^{-3}$. (Taken from Moser et al. 2013). The corresponding first-principles spectra from Verdi et al. (2017) are shown in panels (**d**)–(**f**). The calculated spectral functions were multiplied by the Fermi-Dirac distribution at the experimental temperature ($T = 20$ K) and were convoluted with Gaussian masks of widths 25 meV and 0.015 Å$^{-1}$ in order to account for the experimental resolution in energy and momentum, respectively

are too slow to screen the long-range electric field generated by the $E_u$ phonon vibrations. In this case satellites appear in the spectra, and the electron-phonon coupling strength is approximately independent of doping. When $\omega_P > \omega_{LO}$, in the Fermi liquid regime, the Fröhlich coupling is strongly suppressed, with the polaron satellites gradually replaced by kinks. Correspondingly, the coupling strength decreases. This first-principles analysis indicated that the interplay between lattice vibrations and plasma oscillations can have a strong impact on the polaronic properties of charge carriers in doped oxides.

**Fig. 7** Polaronic and Fermi liquid regimes in $n$-doped anatase $TiO_2$, from Verdi et al. (2017): the red disks indicate the plasma energy at each doping level; the blue disks indicate the electron-phonon coupling strength $\lambda$. The blue line is a guide to the eye, while the red line represents the relation between the plasma energy and the doping density in the homogeneous electron gas. The horizontal line is the energy of the LO $E_u$ phonon of anatase $TiO_2$, 109 meV
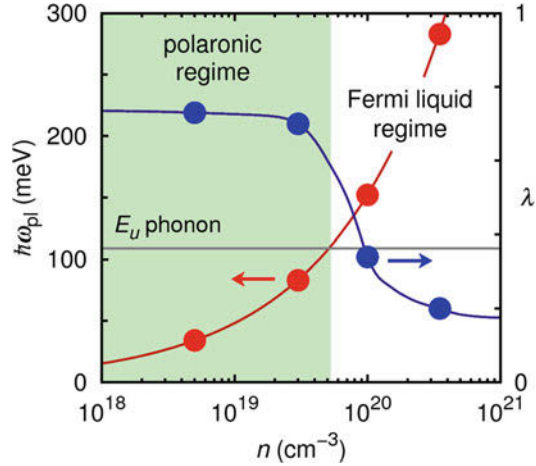


## 6      Hybrid Plasmon-Phonon Satellites

Interestingly, the effects of electron-phonon and electron-plasmon interactions can be readily combined within first-principles calculations if the sum of the relative self-energies (Eqs. (7) and (14)) is included in the calculation of the spectral function, which can thus contain both plasmon and polaron satellite features. This concurrence of plasmon and polaron satellites has been observed experimentally in the case of the ferromagnetic semiconductor EuO and confirmed by first principles calculations (Riley et al. 2018). Experimental ARPES spectra for Gd-doped EuO are reported in Fig. 8a–c for three different doping concentrations, showing the bottom of the conduction band centered at the X point of the Brillouin zone. The energy distribution curves (EDCs) at the conduction band minimum for several dopings are reproduced in Fig. 8g, and they clearly show a shoulder peak whose energy separation with respect to the main quasiparticle band increases with carrier concentration. Such a satellite peak is not resolved above a carrier density $n \approx 10^{20}$ cm$^{-3}$, whereas at low carrier concentration ($n \approx 10^{18}$ cm$^{-3}$) two additional satellites can be distinguished. The spectra calculated with the cumulant expansion method including electron-phonon and electron-plasmon coupling on the same footing are presented in Fig. 8d–f, and they reproduce the features seen in the experiment. In particular, the calculations confirmed that for the lowest doping concentration the series of satellites is mainly due to phonon excitations, with the LO phonon energy of EuO being compatible with the peak separation energy of about 56 meV. Moving to higher dopings, the polar electron-phonon coupling is gradually suppressed by the free-carrier screening, and the satellite peak shifts to higher binding energies. Since the plasma energy increases as the square root of the carrier density, this finding constitutes a fingerprint of the coupling of electrons to plasmonic excitations of the conduction electrons. The first-principles calculations confirmed that the renormalization of the spectral properties at higher dopings is
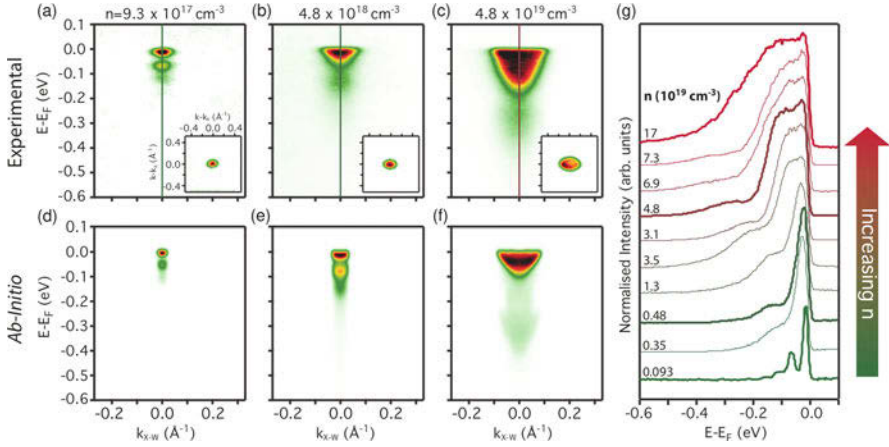
**Fig. 8** (**a**)–(**c**) Measured ARPES spectra of EuO samples with increasing carrier concentrations as indicated on top of each panel, with the Fermi surface contours shown in the insets. The corresponding first-principles data are shown in (**d**)–(**f**). To directly compare with the experiments, the calculated spectral functions were convoluted with two Gaussian masks of widths 20 meV and 0.015 Å$^{-1}$, and integrated along the out-of-plane direction $k_z$. (**g**) Measured energy distribution curves taken at $k = k_X$ (conduction band minimum) for different doping levels. (Figure adapted from Riley et al. 2018)

due to the interplay between electron-phonon and electron-plasmon coupling, and that in particular the low-energy broad satellite seen in Fig. 8c, f is due to plasmonic excitations (Riley et al. 2018).

We remark that the calculations and methods presented so far neglect the effects of mutual renormalization between plasmon and phonon modes, which can arise when the frequency of plasmon and phonon oscillations are of the same order (Varga 1965; Settnes et al. 2017). The inclusion of these effects entirely from first principles represents one of the challenges still open in the investigation of the spectral properties of doped systems.

## 7    Conclusions

The emergence of satellites in photoemission spectroscopy is a universal manifestation of electron-boson interactions in solids. The origin of satellites can be ascribed to the excitation of different types of bosonic modes such as valence plasmons, extrinsic plasmons, or polar phonons. These spectral features have thus far been observed in metals, semiconductors, and highly-doped oxides. Despite the diversity of the physical processes that underpin the satellite formation, and the broad energy scales (from 50–100 meV up to 15–20 eV), many-body perturbation theory provides a unified framework for their description. In combination with standard approximations for the electron-electron and electron-phonon self-energies, the

cumulant expansion approach is a powerful tool for investigating the emergence of spectral fingerprint of electron-boson coupling in solids.

The study of satellites in solids has thus far provided valuable insight into the many-body interactions between electrons, plasmons, and phonons. Recent work in this area has demonstrated that first-principles techniques have reached an accuracy sufficient to even precede experiments in discovering new hallmarks of the coupling between electrons and bosons. The emergence of satellites in photoemission spectroscopy is just one facet of the many effects that electron-boson interaction may induce. The recent findings discussed in this chapter call for a systematic investigation of the influence of low-energy plasmons on the formation of photoemission kinks, waterfall effects, as well as novel mechanisms of superconductive pairing. Furthermore, other spectroscopic techniques, such as absorption, electron energy loss, or time-resolved spectroscopies, provide less explored tools for investigating the coupling between electrons, plasmons, and phonons. Highly doped oxides constitute a particularly exciting playground for exploring the influence of these phenomena on the optoelectronic properties and possible opportunities for exploiting these new emergent properties. In these compounds, the interplay of carriers, extrinsic plasmons, and polar phonons induces complex spectral features that reflect the simultaneous excitation of plasmon and phonon modes and that are highly tunable via the carrier concentration.

In conclusion, the last few years have witnessed a remarkable increase in the accuracy of theoretical techniques for the description of the excited-state phenomena from first principles. These advances, alongside with a relentless increase in experimental resolution, are contributing to strengthen the synergy between theoretical and experimental research, providing numerous opportunities to unveil and understand unexplored forms of fermion-boson coupling in quantum matter.

# References

Allen PB, Heine V (1976) Theory of the temperature dependence of electronic band structures. J Phys C 9:2305

Antonius G, Poncé S, Lantagne-Hurtubise E, Auclair G, Gonze X, Côté M (2015) Dynamical and anharmonic effects on the electron-phonon coupling and the zero-point renormalization of the electronic structure. Phys Rev B 92:085137

Aryasetiawan F, Hedin L, Karlsson K (1996) Multiple plasmon satellites in Na and Al spectral functions from ab initio cumulant expansion. Phys Rev Lett 77:2268–2271

Aryasetiawan F, Gunnarson O (1998) The GW method. Rep Prog Phys 61:237–312

Aulbur WG, Jonsson L, Wilkins JW (2000) Quasiparticle calculations in solids. In: Ehrenreich H, Spaepen F (eds) Solid state physics, vol 54. Academic Press, New York, pp 1–218

Baer Y, Busch G (1973) X-ray photoemission from aluminum. Phys Rev Lett 30:280–282

Bostwick A, Speck F, Seyller T, Horn K, Polini M, Asgari R, MacDonald AH, Rotenberg E (2010) Observation of plasmarons in quasi-freestanding doped graphene. Science 328:999–1002

Cancellieri C, Mishchenko AS, Aschauer U, Filippetti A, Faber C, Barišić OS, Rogalev VA, Schmitt T, Nagaosa N, Strocov VN (2016) Polaronic metal state at the LaAlO$_3$/SrTiO$_3$ interface. Nat Commun 7:10386

Caruso F, Giustino F (2015) Spectral fingerprints of electron-plasmon coupling. Phys Rev B 92:045123

Caruso F, Giustino F (2016) Theory of electron-plasmon coupling in semiconductors. Phys Rev B 94:115208

Caruso F, Rinke P, Ren X, Rubio A, Scheffler M (2013) Self-consistent $GW$: all-electron implementation with localized basis functions. Phys Rev B 88:075105

Caruso F, Lambert H, Giustino F (2015) Band structures of plasmonic polarons. Phys Rev Lett 114:146404

Caruso F, Dauth M, van Setten MJ, Rinke P (2016) Benchmark of $GW$ approaches for the $GW$100 test set. J Chem Theory Comput 12:5076–5087

Caruso F, Verdi C, Poncé S, Giustino F (2018) Electron-plasmon and electron-phonon satellites in the angle-resolved photoelectron spectra of n-doped anatase TiO2. Phys Rev B 97:165113

Chang YJ, Bostwick A, Kim YS, Horn K, Rotenberg E (2010) Structure and correlation effects in semiconducting $SrTiO_3$. Phys Rev B 81:235109

Chen C, Avila J, Frantzeskakis E, Levy A, Asensio MC (2015) Observation of a two-dimensional liquid of Fröhlich polarons at the bare $SrTiO_3$ surface. Nat Commun 6:8585

Damascelli A, Hussain Z, Shen ZX (2003) Angle-resolved photoemission studies of the cuprate superconductors. Rev Mod Phys 75:473

Eiguren A, Hellsing B, Reinert F, Nicolay G, Chulkov EV, Silkin VM, Hüfner S, Echenique PM (2002) Role of bulk and surface phonons in the decay of metal surface states. Phys Rev Lett 88:066805

Fröhlich H (1954) Electrons in lattice fields. Adv Phys 3:325

Giustino F (2017) Electron-phonon interactions from first principles. Rev Mod Phys 89:015003

Giustino F, Cohen ML, Louie SG (2007) Electron-phonon interaction using Wannier functions. Phys Rev B 76:165108

Giustino F, Louie SG, Cohen ML (2010) Electron-phonon renormalization of the direct band gap of diamond. Phys Rev Lett 105:265501

Gumhalter B, Kovač V, Caruso F, Lambert H, Giustino F (2016) On the combined use of $GW$ approximation and cumulant expansion in the calculations of quasiparticle spectra: the paradigm of Si valence bands. Phys Rev B 94:035103

Guzzo M, Lani G, Sottile F, Romaniello P, Gatti M, Kas JJ, Rehr JJ, Silly MG, Sirotti F, Reining L (2011) Valence electron photoemission spectrum of semiconductors: ab initio description of multiple satellites. Phys Rev Lett 107:166401

Guzzo M, Kas JJ, Sottile F, Silly MG, Sirotti F, Rehr JJ, Reining L (2012) Plasmon satellites in valence-band photoemission spectroscopy. Eur Phys J B 85:324

Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139:A796

Hedin L (1980) Effects of recoil on shake-up spectra in metals. Phys Scr 21:477

Hedin L, Lundqvist S (1970) Effects of electron-electron and electron-phonon interactions on the one-electron states of solids. Solid State Phys 23:1–181

Holm B, Aryasetiawan F (1997) Self-consistent cumulant expansion for the electron gas. Phys Rev B 56:12825–12831

Holm B, von Barth U (1998) Fully self-consistent $GW$ self-energy of the electron gas. Phys Rev B 57:2108

Hüfner S (2003) Photoelectron spectroscopy, 3rd edn. Springer, Berlin

Kas JJ, Rehr JJ, Reining L (2014) Cumulant expansion of the retarded one-electron Green function. Phys Rev B 90:085112

King PDC, McKeown Walker S, Tamai A, de la Torre A, Eknapakul T, Buaphet P, Mo SK, Meevasana W, Bahramy MS, Baumberger F (2014) Quasiparticle dynamics and spin-orbital texture of the $SrTiO_3$ two-dimensional electron gas. Nat Commun 5:3414

Kittel C (1976) Introduction to solid state physics, 5th edn. Wiley, New York

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://link.aps.org/doi/10.1103/PhysRev.140.A1133

Kutepov A, Haule K, Savrasov SY, Kotliar G (2012) Electronic structure of Pu and Am metals by self-consistent relativistic $GW$ method. Phys Rev B 85:155129

Langreth DC (1970) Singularities in the X-ray spectra of metals. Phys Rev B 1:471–477

Lanzara A, Bogdanov PV, Zhou XJ, Kellar SA, Feng DL, Lu ED, Yoshida T, Eisaki H, Fujimori A, Kishio K, Shimoyama JI, Noda T, Uchida S, Hussain Z, Shen ZX (2001) Evidence for ubiquitous strong electron-phonon coupling in high-temperature superconductors. Nature 412:510

Lee JJ, Schmitt FT, Moore RG, Johnston S, Cui YT, Li W, Yi M, Liu ZK, Hashimoto M, Zhang Y, Lu DH, Devereaux TP, Lee DH, Shen ZX (2014) Interfacial mode coupling as the origin of the enhancement of $T_c$ in FeSe films on $SrTiO_3$. Nature 515:245–248

Lischner J, Pálsson GK, Vigil-Fowler D, Nemsak S, Avila J, Asensio MC, Fadley CS, Louie SG (2015) Satellite band structure in silicon caused by electron-plasmon coupling. Phys Rev B 91:205113

Logothetidis S, Petalas J, Polatoglou HM, Fuchs D (1992) Origin and temperature dependence of the first direct gap of diamond. Phys Rev B 46:4483–4494

Lundqvist BI (1967) Single-particle spectrum of the degenerate electron gas. Phys Kondens Mater 6:193–205

Mahan G (2000) Many-particle physics. Springer, New York

Marini A, Hogan C, Grning M, Varsano D (2009) Yambo: an ab initio tool for excited state calculations. Comput Phys Commun 180:1392–1403

Marini A, Poncé S, Gonze X (2015) Many-body perturbation theory approach to the electron-phonon interaction with density-functional theory as a starting point. Phys Rev B 91:224310

Moser S, Moreschini L, Jaćimović J, Barišić OS, Berger H, Magrez A, Chang YJ, Kim KS, Bostwick A, Rotenberg E, Forró L, Grioni M (2013) Tunable polaronic conduction in anatase $TiO_2$. Phys Rev Lett 110:196403

Nery JP, Allen PB, Antonius G, Reining L, Miglio A, Gonze X (2018) Quasiparticles and phonon satellites in spectral functions of semiconductors and insulators: cumulants applied to the full first-principles theory and the Fröhlich polaron. Phys Rev B 97:115145

Park CH, Giustino F, Cohen ML, Louie SG (2007) Velocity renormalization and carrier lifetime in graphene from the electron-phonon interaction. Phys Rev Lett 99:086804

Poncé S, Gillet Y, Laflamme Janssen J, Marini A, Verstraete M, Gonze X (2015) Temperature dependence of the electronic structure of semiconductors and insulators. J Chem Phys 143:102813

Poncé S, Margine ER, Verdi C, Giustino F (2016) EPW: electron-phonon coupling, transport and superconducting properties using maximally localized Wannier functions. Comput Phys Commun 209:116–133

Rademaker L, Wang Y, Berlijn T, Johnston S (2016) Enhanced superconductivity due to forward scattering in fese thin films on $SrTiO_3$ substrates. New J Phys 18:022001

Riley JM, Caruso F, Verdi C, Duffy LB, Watson MD, Bawden L, Volckaert K, van der Laan G, Hesjedal T, Hoesch M, Giustino F, King PDC (2018) Crossover from lattice to plasmonic polarons of a spin-polarised electron gas in ferromagnetic EuO. Nat Commun 9:2305

Settnes M, Saavedra JRM, Thygesen KS, Jauho AP, Garca de Abajo FJ, Mortensen NA (2017) Strong plasmon-phonon splitting and hybridization in 2D materials revealed through a self-energy approach. ACS Photon 4(11):2908–2915

Sjakste J, Vast N, Calandra M, Mauri F (2015) Wannier interpolation of the electron-phonon matrix elements in polar semiconductors: polar-optical coupling in GaAs. Phys Rev B 92:054307

Steiner P, Höchst H, Hüfner S (1979) Photoemission in solids II. In: Ley L, Cardona M (eds) Topics in applied physics, vol 27. Springer, Heidelberg

Story SM, Kas JJ, Vila FD, Verstraete MJ, Rehr JJ (2014) Cumulant expansion for phonon contributions to the electron spectral function. Phys Rev B 90:195135

Varga BB (1965) Coupling of plasmons to polar phonons in degenerate semiconductors. Phys Rev 137:A1896

Verdi C, Giustino F (2015) Fröhlich electron-phonon vertex from first principles. Phys Rev Lett 115:176401

Verdi C, Caruso F, Giustino F (2017) Origin of the crossover from polarons to Fermi liquids in transition metal oxides. Nat Commun 8:15769

Wang Z, McKeown Walker S, Tamai A, Wang Y, Ristic Z, Bruno FY, de la Torre A, Riccò S, Plumb NC, Shi M, Hlawenka P, Sánchez-Barriga J, Varykhalov A, Kim TK, Hoesch M, King PDC, Meevasana W, Diebold U, Mesot J, Moritz B, Devereaux TP, Radovic M, Baumberger F (2016) Tailoring the nature and strength of electron-phonon interactions in the $SrTiO_3(001)$ two-dimensional electron liquid. Nat Mater 15:835–839

Yukawa R, Ozawa K, Yamamoto S, Iwasawa H, Shimada K, Schwier EF, Yoshimatsu K, Kumigashira H, Namatame H, Taniguchi M, Matsuda I (2016) Phonon-dressed two-dimensional carriers on the ZnO surface. Phys Rev B 94:165313

Zhou JS, Kas JJ, Sponza L, Reshetnyak I, Guzzo M, Giorgetti C, Gatti M, Sottile F, Rehr JJ, Reining L (2015) Dynamical effects in electron spectroscopy. J Chem Phys 143:184109

# Non-equilibrium Green's Functions for Coupled Fermion-Boson Systems

**17**

Daniel Karlsson  and Robert van Leeuwen

## Contents

### Abstract

In this book section, we give a brief introduction to many-body perturbation theory for coupled fermion-boson systems using non-equilibrium Green's functions. Using the language of modern many-body perturbation theory and the so-called contour-ordered correlators, a single consistent formalism arises which can be applied to a multitude of classes of systems.

D. Karlsson · R. van Leeuwen (✉)
Department of Physics, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland
e-mail: daniel.l.e.karlsson@jyu.fi; robert.vanleeuwen@jyu.fi

# 1    Introduction

The many-body problem is at the heart of the field of condensed matter physics. While the Schrödinger equation is impossible to solve except for the smallest of systems, condensed matter physics deal with the whole spectrum of few-body to many-body physics. This makes it difficult to consider the many-body wave function as a useful object.

Another different route is to try to lower one's ambitions. While the wave function gives access to all measurable observables in a system, including high-order correlation functions, experimentally one is most often interested in low-order correlation functions such as one-particle correlation functions (e.g., particle densities and electric currents) or two-particle correlation functions (e.g., polarizabilities and magnetic susceptibilities). As such, it has been useful to concentrate on obtaining approximate expressions for reduced objects.

Non-equilibrium Green's functions (NEGF) offer one such route. The one-particle Green's function contains all information about one-particle observables, and similarly higher-order Green's functions contain information on higher-order correlators. The complexity of these objects is much lower than for the full wave function, and powerful diagrammatic techniques for calculating the Green's function as a perturbation series are available. Moreover, the NEGF formalism unlocks the door to out-of-equilibrium physics, such as transient effects in quantum transport and pump-probe spectroscopy.

There is plenty of literature concerning different types of electron-boson interactions treated within many-body perturbation theory in (e.g., Fetter and Walecka 2003; Bruus and Flensberg 2004) and out (e.g., Henneberger et al. 2000; Bonitz 2016) of equilibrium. The theoretical structure for different kind of bosons is very similar, as also stressed in Schüler et al. (2016), where quantized photons, phonons, and other excitations can be treated in an almost identical way. They give the example of time-dependent photoemission under the influence of plasmonic excitations. Other electron-boson systems are, for example, electron-magnon systems (Mahfouzi and Nikolić 2014) and electron-phonon systems (Sentef et al. 2016).

We begin with a general introduction to NEGF, and diagrammatic perturbation theory will be systematically built up. The discussion will follow the notation and structure of Stefanucci and van Leeuwen (2013).

# 2    Introduction to NEGF

## 2.1    Time-Dependent Observables

We consider a system containing two different species of particles: fermions and bosons. The system is initially prepared in an ensemble at time $t = t_0$, described by the statistical density matrix $\hat{\rho}(t_0) = \hat{\rho}$. The initial ensemble average $O(t_0)$ of a

general operator $\hat{O}(t)$ is defined as $O(t_0) = \mathrm{Tr}\left[\hat{\rho}\hat{O}(t_0)\right]$, where the trace $\mathrm{Tr}[\cdots]$ is over all states and particle numbers in Fock space. We write the density matrix as $\hat{\rho} = e^{-\beta\hat{H}^M}/Z$, where the Matsubara Hamiltonian $\hat{H}^M = \hat{H} - \mu\hat{N}$ determines the initial state, where $\hat{H} = \hat{H}(t_0)$ is the equilibrium Hamiltonian. $\mu$ is the chemical potential for the fermions, $\hat{N}$ is the corresponding particle number operator, and $\beta$ is the inverse temperature. The partition function $Z = \mathrm{Tr}\left[e^{-\beta\hat{H}^M}\right]$ yields $\mathrm{Tr}\left[\hat{\rho}\right] = 1$.

For times $t > t_0$, we imagine that the system is perturbed to bring it out of equilibrium. This is described by the time-dependent Hamiltonian $\hat{H}(t)$, under which the density matrix changes according to

$$\hat{\rho}(t) = \hat{U}(t, t_0)\hat{\rho}\hat{U}(t_0, t). \tag{1}$$

We introduced the time-evolution operator $U(t, t_0)$, defined as

$$\hat{U}(t_1, t_2) = \begin{cases} \mathscr{T}\left\{e^{-i\int_{t_1}^{t_2} d\bar{t}\,\hat{H}(\bar{t})}\right\} & \text{for } t_2 \text{ later than } t_1 \\ \bar{\mathscr{T}}\left\{e^{+i\int_{t_2}^{t_1} d\bar{t}\,\hat{H}(\bar{t})}\right\} & \text{for } t_1 \text{ later than } t_2 \end{cases} \tag{2}$$

where $\mathscr{T}\{\cdots\}$ ($\bar{\mathscr{T}}\{\cdots\}$) denotes (anti)time-ordering, which orders the latest (earliest) time to the left.

Using the cyclic property of the trace, we can write the time-dependent ensemble average $O(t) = \mathrm{Tr}\left[\hat{\rho}(t)\hat{O}(t)\right]$ as

$$O(t) = \mathrm{Tr}\left[\hat{\rho}\hat{U}(t_0, t)\hat{O}(t)\hat{U}(t, t_0)\right]. \tag{3}$$

Using the definitions, we can write Eq. (3) as

$$O(t) = \mathrm{Tr}\left[\hat{\rho}\bar{\mathscr{T}}\left\{e^{-i\int_t^{t_0} d\bar{t}\hat{H}(\bar{t})}\right\}\hat{O}(t)\mathscr{T}\left\{e^{-i\int_{t_0}^t d\bar{t}\hat{H}(\bar{t})}\right\}\right]. \tag{4}$$

Equation (4) is a bit awkward to handle since it contains two different types of time-ordering operators. The ensemble average can be written in a more convenient way by introducing a vertical track via $e^{-\beta\hat{H}^M} = e^{-i\int_{t_0}^{t_0-i\beta} d\bar{z}\hat{H}^M(\bar{z})}$, where $\hat{H}^M(z) = \hat{H}^M$ is constant. The ensemble average of Eq. (4) can then be viewed as three different time-evolution operators, one going forward in time from $t_0$ to $t$, then going backward in time from $t$ to $t_0$, and then finally going from $t_0$ to $t_0 - i\beta$. All these different time-evolution operators can be written using a single time-ordering by introducing a directed time-ordered contour $\gamma$, depicted in Fig. 1. The contour consists of three branches. Times on the forward and backward branches are denoted $t^-$ and $t^+$, respectively, while times on the vertical (Matsubara) branch are denoted $t^M$. A general contour time is denoted by $z$, and we define $\hat{H}(z = $
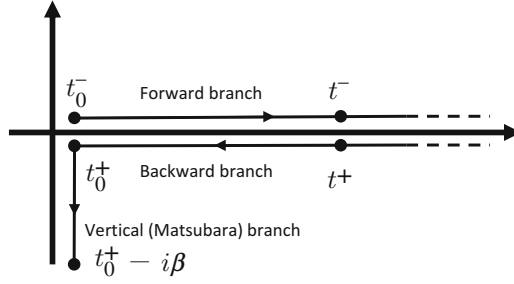
**Fig. 1** The contour. The equations of motion are defined on this contour, consisting of a forward branch, backward branch, and an imaginary (Matsubara) branch. The initial point on the contour is $z_i = t_0^-$, and the final point is $z_f = t_0^+ - i\beta$. The contour can be deformed into various other ones used in the literature

$t^\pm) = \hat{H}(t)$, $\hat{H}(z = t^M) = \hat{H}^M$. For all other operators we consider, we define $\hat{O}(z = t^\pm) = \hat{O}(t)$, $\hat{O}(z = t^M) = \hat{O}(t_0)$. The use of the contour allows us to write the ensemble average of Eq. (4) in the elegant form

$$O(z) = \frac{\text{Tr}\left[\mathscr{T}_c\left\{e^{-i\int_\gamma d\bar{z}\,\hat{H}(\bar{z})}\hat{O}(z)\right\}\right]}{\text{Tr}\left[\mathscr{T}_c\left\{e^{-i\int_\gamma d\bar{z}\,\hat{H}(\bar{z})}\right\}\right]}. \tag{5}$$

where $\mathscr{T}_c\{\cdots\}$ denotes contour time-ordering, in the sense of the directed contour. Earlier contour times appear first when following the directed contour starting from $t_0^-$. For an in-depth discussion, see, for example, Stefanucci and van Leeuwen (2013) and Špička et al. (2014).

The time contour allows for elegant expressions of non-equilibrium physics via Eq. (5). Furthermore, the contour yields simple equations of motion for Heisenberg operators. We define an arbitrary Heisenberg operator $\hat{O}_H(z)$ as

$$\hat{O}_H(z) = \hat{U}(z_i, z)\hat{O}(z)\hat{U}(z, z_i) \tag{6}$$

where $z_i$ is the initial time on the contour, and the time-evolution operator on the contour is given by

$$\hat{U}(z_2, z_1) = \begin{cases} \mathscr{T}_c\left\{e^{-i\int_{z_1}^{z_2} d\bar{z}\,\hat{H}(\bar{z})}\right\} & \text{for } z_2 \text{ later than } z_1 \\ \bar{\mathscr{T}}_c\left\{e^{+i\int_{z_2}^{z_1} d\bar{z}\,\hat{H}(\bar{z})}\right\} & \text{for } z_1 \text{ later than } z_2 \end{cases} \tag{7}$$

where $\bar{\mathscr{T}}_c\{\cdots\}$ denotes contour anti-time-ordering. Using the time-evolution operator defined on the contour, we can derive the contour Heisenberg equation of motion:

$$i \frac{\mathrm{d}}{\mathrm{d}z} \hat{O}_H(z) = \left[ \hat{O}_H(z), \hat{H}_H(z) \right]_- + \hat{U}(z_i, z) \left( i \frac{\mathrm{d}}{\mathrm{d}z} \hat{O}(z) \right) \hat{U}(z, z_i) \qquad (8)$$

These are the basic equations necessary to describe systems out of equilibrium. We will now specify the type of systems we will study.

## 2.2   The Structure of the Hamiltonian

We consider a time-dependent fermion-boson Hamiltonian as

$$\hat{H}(z) = \underbrace{\hat{H}_{\mathrm{el}}(z) + \hat{H}_{\mathrm{bos}}(z)}_{\hat{H}_0(z)} + \underbrace{\hat{H}_{\mathrm{el\text{-}el}}(z) + \hat{H}_{\mathrm{el\text{-}bos}}(z)}_{\hat{H}_{\mathrm{int}}(z)}, \qquad (9)$$

where the non-interacting part $\hat{H}_0$ consists of one-body operators and the interaction part $\hat{H}_{\mathrm{int}}$ consists of higher-order operators. To avoid confusion and to increase clarity, we write that the two different kinds of particles are electrons and bosons. We note, however, that the same formalism will be valid if $\hat{H}_{\mathrm{el}}$ and $\hat{H}_{\mathrm{el\text{-}el}}$ would refer to general fermions or bosons. The difference is in the (anti)commutation relations, which we denote $\left[ \hat{O}_1, \hat{O}_2 \right]_{\mp} = \hat{O}_1 \hat{O}_2 \mp \hat{O}_2 \hat{O}_1$, where the upper/lower sign is for bosons/fermions.

The purely electronic part is

$$\hat{H}_{el}(z) = \int \mathrm{d}x \mathrm{d}x' \hat{\psi}^\dagger(x) \left\langle x \left| \hat{h}(z) \right| x' \right\rangle \hat{\psi}(x'), \qquad (10)$$

where $x$ denotes a collective space-spin variable. The field creation operator $\hat{\psi}^\dagger$ and destruction operator $\hat{\psi}$ obey the usual (anti)commutation relations $\left[ \hat{\psi}(x), \hat{\psi}(y) \right]_{\mp} = \left[ \hat{\psi}^\dagger(x), \hat{\psi}^\dagger(y) \right]_{\mp} = 0$, $\left[ \hat{\psi}(x), \hat{\psi}^\dagger(y) \right]_{\mp} = \delta(x - y)$. To simplify the notation, we will assume $\left\langle x_1 \left| \hat{h}(z_1) \right| x_2 \right\rangle = \delta(x_1 - x_2) h(x_1, z_1)$, that is, that the one-body part does not contain spin-flip terms. The one-body term $h(x, z)$ contains the kinetic energy and a general time-dependent external potential.

The electrons interact among each other via a two-body interaction $\hat{H}_{\mathrm{el\text{-}el}}$:

$$\hat{H}_{\mathrm{el\text{-}el}}(z) = \frac{1}{2} \int \mathrm{d}x \mathrm{d}x' \hat{\psi}^\dagger(x) \hat{\psi}^\dagger(x') v(x, x', z) \hat{\psi}(x') \hat{\psi}(x), \qquad (11)$$

where $v(x, x', z)$ is the time-dependent pair interaction. The interaction can be coulombic, but we allow for more freedom by allowing a general two-body interaction; this allows for considering, for example, local interactions where $v(x, x', z) \sim \delta(x - x')$, ubiquitous in model systems and in cold atom gases (Bloch and Zwerger 2008). We also allow for a time-dependence that permits to describe, for example, interaction quenches or an adiabatic switch-on of the interaction.

We now focus on the bosons, which we will treat slightly differently. The fundamental reason for this is that the interaction part of the electron-boson Hamiltonian is not particle conserving. Let us first consider the noninteracting part $\hat{H}_{\text{bos}}$ as

$$\hat{H}_{\text{bos}}(z) = \sum_\mu \omega_\mu(z) \hat{a}_\mu^\dagger \hat{a}_\mu, \tag{12}$$

in the diagonal form where $\omega_\mu(z)$ is the frequency of mode $\mu$ of the bosons and $\hat{a}_\mu^\dagger (\hat{a}_\mu)$ are the corresponding creation (annihilation) operator for mode $\mu$. The operators fulfill the usual commutation relations $[\hat{a}_\mu, \hat{a}_\nu] = [\hat{a}_\mu^\dagger, \hat{a}_\nu^\dagger] = 0$ and $[\hat{a}_\mu, \hat{a}_\nu^\dagger] = \delta_{\mu\nu}$. There are many situations in which it is more convenient to work with the Hermitian displacement $\hat{u}_\mu$ and momentum $\hat{p}_\mu$ operators, defined via the transformation

$$\hat{u}_\mu = \left(\hat{a}_\mu^\dagger + \hat{a}_\mu\right)/\sqrt{2}, \quad \hat{p}_\mu = i\left(\hat{a}_\mu^\dagger - \hat{a}_\mu\right)/\sqrt{2}. \tag{13}$$

The displacement and momentum operators fulfill the usual commutation relations $[\hat{u}_\mu, \hat{u}_\nu] = [\hat{p}_\mu, \hat{p}_\nu] = 0$, and $[\hat{u}_\mu, \hat{p}_\nu] = i\delta_{\mu\nu}$. In order to prepare the discussion for more general situations, it turns out that a convenient description can be obtained if the $\hat{u}_\mu$ and $\hat{p}_\mu$ operators are gathered into a vector $\hat{\phi}_{\mu\xi}$, with $\hat{\phi}_{\mu,1} = \hat{u}_\mu$ and $\hat{\phi}_{\mu,2} = \hat{p}_\mu$. We also introduce the composite index $\bar{\mu} = (\mu, \xi_\mu)$ to simplify the notation. The bosonic Hamiltonian, Eq. (12), can be generalized to

$$\hat{H}_{\text{bos}}(z) = \sum_{\bar{\mu}\bar{\nu}} \Omega_{\bar{\mu}\bar{\nu}}(z) \hat{\phi}_{\bar{\mu}} \hat{\phi}_{\bar{\nu}}. \tag{14}$$

Equation (14) is then written using Hermitian operators, which generalize Eq. (12). More specifically, Eq. (14) can contain terms proportional to $(\hat{a}^\dagger)^2$ and $\hat{a}^2$, thereby violating particle number conservation. The commutation relations between the $\hat{\phi}$-fields can be given compactly as

$$\left[\hat{\phi}_{\bar{\mu}}, \hat{\phi}_{\bar{\nu}}\right]_- = \alpha_{\bar{\mu}\bar{\nu}}, \quad \text{with} \quad \alpha_{\bar{\mu}\bar{\nu}} = \delta_{\mu\nu} \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}_{\xi_\mu \xi_\nu}. \tag{15}$$

The advantage of the $\hat{\phi}$-field vector formulation will become clear when we later discuss the equations of motion. It implies that instead of using second-order equations for the displacement, we can use a first-order equation for the coupled displacement and momentum operators.

In order to include the multitude of electron-boson interactions in the literature, we take the interaction to be

$$\hat{H}_{\text{el-bos}}(z) = \sum_{\bar{\mu}} \int dx \ \hat{\psi}^{\dagger}(x)\lambda_{\bar{\mu}}(x,z)\hat{\psi}(x)\hat{\phi}_{\bar{\mu}}. \tag{16}$$

As in the case of $v(x, x', z)$, the (real-valued) coupling strength $\lambda_{\bar{\mu}}(x,z)$ can be time-dependent. The electron-boson interaction could be written more generally by letting $\lambda_{\bar{\mu}}(x,z)$ being a differential operator, and this will not change the overall discussion. However, for simplicity, we let $\lambda_{\bar{\mu}}(x,z)$ be a local operator. Equivalent forms of the electron-boson interaction can be obtained by defining $m_{\mu}(x) = (\lambda_{\mu,1}(x) - i\lambda_{\mu,2}(x))/\sqrt{2}$, which yields

$$\hat{H}_{\text{el-bos}} = \sum_{\mu} \int dx \ \hat{\psi}^{\dagger}(x)\hat{\psi}(x)\left(m_{\mu}(x)\hat{a}_{\mu}^{\dagger} + m_{\mu}^{*}(x)\hat{a}_{\mu}\right).$$

Another form can be obtained by choosing a discrete electron basis, $\hat{\psi}(x) = \sum_{l} \varphi_{l}(x)\hat{c}_{l}$, which yields $\hat{H}_{\text{el-bos}} = \sum_{\mu kl} \hat{c}_{k}^{\dagger}\hat{c}_{l}(m_{\mu}^{kl}\hat{a}_{\mu}^{\dagger} + m_{\mu}^{lk,*}\hat{a}_{\mu})$, where the elements $m_{\mu}^{kl} = \int dx \ \varphi_{k}^{*}(x)m_{\mu}(x)\varphi_{l}(x)$. Written in the above way, the general interaction Hamiltonian, Eq. (16), includes the commonly used electron-phonon interaction, electron-plasmon, electron-magnon, electron-photon in cavity quantum electrodynamics, and others.

We finally note that we have not included the so-called half-body terms in the bosonic Hamiltonian, terms proportional to a single $\hat{\phi}$ operator. These terms are important in several contexts, such as in the description of forced oscillators and superfluid systems, but the half-body interactions can change the diagrammatic structure, for example, in a Bose-Einstein condensate. For a thorough diagrammatic treatment of half-body interactions, see De Dominicis (1963). Keeping this in mind, we will limit ourselves to the Hamiltonian of Eq. (9).

## 2.3 Operator Correlators

To calculate time-dependent ensemble averages, like in Eq. (5), we need to deal with time-ordered products of arbitrary operators – operator correlators. In this section, we derive general equations of motion for such objects. The time-ordering operator for the same particle species acts in the following way:

$$\mathcal{T}_{c}\left\{\hat{O}_{1}(z_{1})\hat{O}_{2}(z_{2})\right\} = \theta(z_{1}, z_{2})\hat{O}_{1}(z_{1})\hat{O}_{2}(z_{2}) \pm \theta(z_{2}, z_{1})\hat{O}_{2}(z_{2})\hat{O}_{1}(z_{1}), \tag{17}$$

that is, the time-ordering orders the later contour time to the left. Here, as in the rest of the chapter, the upper/lower sign refers to bosonic/fermionic operators. The functions $\theta(z_{1}, z_{2})$ are step functions on the contour, with $\theta(z_{1}, z_{2}) = 1$ if $z_{1}$ is later than $z_{2}$ on the contour and 0 otherwise. We define the Dirac delta function on the contour as $\delta(z_{1}, z_{2}) = \partial_{z_{1}}\theta(z_{1}, z_{2}) = -\partial_{z_{2}}\theta(z_{1}, z_{2})$. For a string of operators

that contain operators pertaining to different species of particles, we define them to commute under the time-ordering operator.

The equations of motion for the time-ordered product is given by

$$\frac{\partial}{dz_1} \mathcal{T}_c \left\{ \hat{O}_1(z_1) \hat{O}_2(z_2) \right\} = \delta(z_1, z_2) \left[ \hat{O}_1, \hat{O}_2 \right]_{\mp} + \mathcal{T}_c \left\{ \frac{\partial}{\partial z_1} \hat{O}_1(z_1) \hat{O}_2(z_2) \right\}. \tag{18}$$

The generalization of Eq. (18) to an arbitrary correlator is

$$\frac{\partial}{\partial z_k} \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{O}_n \right\} = \partial^{\theta}_{z_k} \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{O}_n \right\}$$
$$+ \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{O}_{k-1} \left( \frac{\partial}{\partial z_k} \hat{O}_k \right) \hat{O}_{k+1} \dots \hat{O}_n \right\} \tag{19}$$

where we have defined the contribution from differentiating the step functions in the first term on the right-hand side. Most often, one is interested in the equations of motion when the operators $\hat{O}_1, \cdots, \hat{O}_n$ fulfill the extra condition that their equal-time (anti)commutators commute with other operators:

$$\left[ \hat{O}_k(z), \hat{O}_l(z) \right]_{\mp} = c_{kl} \hat{\mathbb{1}}, \tag{20}$$

where $c_{kl}$ is a number. We will consider such cases here; the creation and destruction operators, as well as the $\hat{\phi}$-fields, fulfill the relation of Eq. (20). In that case, the explicit form of $\partial^{\theta}_{z_k} \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{O}_n \right\}$ is given by

$$\partial^{\theta}_{z_k} \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{O}_n \right\} = \sum_{l=1}^{k-1} (\pm)^{k-l} \delta(z_k, z_l) \left[ \hat{O}_k, \hat{O}_l \right]_{\mp} \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{\hat{O}}_l \dots \hat{\hat{O}}_k \dots \hat{O}_n \right\}$$
$$+ \sum_{l=k+1}^{n} (\pm)^{l-k-1} \delta(z_k, z_l) \left[ \hat{O}_k, \hat{O}_l \right]_{\mp} \mathcal{T}_c \left\{ \hat{O}_1 \dots \hat{\hat{O}}_k \dots \hat{\hat{O}}_l \dots \hat{O}_n \right\}, \tag{21}$$

where $\hat{\hat{O}}_l$ denotes that the operator $\hat{O}_l$ is missing from the string.

The equation of motion for operator correlators, Eq. (19), is very general, and is the equation that will give rise to the equations of motion for different kinds of Green's functions with different types of particles. To proceed, we need to specialize our discussion to which types of correlators we want to consider.

### 2.3.1  $n$-Particle Correlators on the Contour

In systems that conserve the number of electrons, all observables can be expressed as linear combinations of field operators that contain an equal number of creation and destruction operators. We thus define the $n$-particle correlator $\hat{G}_n$ as

$$\hat{G}_n(1, \ldots, n; 1', \ldots, n') = (-i)^n \mathscr{T}_c \left\{ \hat{\psi}_H(1) \cdots \hat{\psi}_H(n) \hat{\psi}_H^\dagger(n') \cdots \hat{\psi}_H^\dagger(1') \right\}$$

(22)

where unprimed variables $k = (x_k, z_k)$ pertain to destruction operators $\hat{\psi}_H(k)$, while primed variables $k' = (x_k', z_k')$ pertain to creation operators $\hat{\psi}_H^\dagger(k')$. The $2n$-point correlators $\hat{G}_n(1, \ldots, n; 1', \ldots, n')$ fulfill the following equations of motion, which follow from Eqs. (19) and (21) and the Heisenberg equation of motion, Eq. (8):

$$\left[ i \frac{\partial}{\partial z_k} - h(k) \right] \hat{G}_n(1, \ldots, n; 1', \ldots, n')$$

$$= (-i)^n \mathscr{T}_c \left\{ \hat{\psi}_H(1) \ldots \left[ \hat{\psi}_H(k), \hat{H}_{\text{int}}(z_k) \right]_- \ldots \hat{\psi}_H(n) \hat{\psi}_H^\dagger(n') \ldots \hat{\psi}_H^\dagger(1') \right\}$$

$$+ \sum_{j=1}^n (\pm)^{k+j} \delta(k; j') \hat{G}_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, n; 1', \ldots, \overset{\sqcap}{j'}, \ldots, n'),$$

(23)

and likewise for the primed coordinates. We define $\hat{G}_0 = \hat{\mathbb{1}}$.

We now consider the situation where the operators in the correlator are taken to be $\hat{\phi}$-fields. Similarly to the electronic case, we define the $n$-body bosonic correlator:

$$\hat{\mathbb{D}}_n(1, 2, \ldots, 2n) = (-i)^n \mathscr{T}_c \left\{ \hat{\phi}_H(1) \hat{\phi}_H(2) \ldots \hat{\phi}_H(2n) \right\},$$

(24)

where $1 = (\bar{\mu}_1, z_1)$ is a collective index. Here, $n$ can also be a half-integer, because, as we will see, the ensemble average of an odd number of $\hat{\phi}$-fields is nonvanishing when the electron-boson interaction is present.

The equations of motion for $\hat{\mathbb{D}}_n$ can be found in the same way as for $\hat{G}_n$, yielding

$$i \frac{\partial}{\partial z_k} \hat{\mathbb{D}}_n(1, \ldots \bar{\sigma} z_k \ldots, 2n) - \sum_{\bar{\mu}\bar{\nu}} \alpha_{\bar{\sigma}\bar{\mu}} \bar{\Omega}_{\bar{\mu}\bar{\nu}}(z_k) \hat{\mathbb{D}}_n(1, \ldots, \nu z_k, \ldots, 2n)$$

$$= (-i)^n \mathscr{T}_c \left\{ \hat{\phi}_H(1) \ldots \left[ \hat{\phi}_H(k), \hat{H}_{\text{int}}(z_k) \right]_- \ldots \hat{\phi}_H(2n) \right\}$$

$$+ \sum_{j=1, j\neq k}^{2n} \delta(z_k, z_j) \alpha_{\bar{\sigma}, \bar{\mu}_j} \hat{\mathbb{D}}_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, \overset{\sqcap}{j}, \ldots, 2n).$$

(25)

We define $\hat{\mathbb{D}}_0 = \hat{\mathbb{1}}$ and $\hat{\mathbb{D}}_{-1/2} = \hat{0}$. Equation (25) is somewhat different than the one for $\hat{G}_n$, Eq. (23), owing to the presence of the $\alpha$ matrix from the commutation relations. However, the main structure is identical.

As an illustration of Eq. (25), we consider the case of $n = 1/2$. In this case the correlator is the $\hat{\phi}$-field itself, $\hat{\mathbb{D}}_{1/2}(1) = (-i)^{1/2}\hat{\phi}(1)$. Using the commutators $\left[\hat{\phi}_{\bar{\sigma}}, \hat{H}_{\text{el-bos}}\right]_- = \sum_{\bar{\mu}} \int dx \lambda_{\bar{\mu}}(x, z) \alpha_{\bar{\sigma}\bar{\mu}} \hat{\psi}^\dagger(x) \hat{\psi}(x)$ and $\left[\hat{\phi}_{\bar{\sigma}}, \hat{H}_{\text{bos}}\right]_- = \sum_{\bar{\mu}\bar{\nu}} \bar{\Omega}_{\bar{\mu}\bar{\nu}} \alpha_{\bar{\sigma}\bar{\mu}} \hat{\phi}_{\bar{\nu}}$ where $\bar{\Omega}_{\bar{\mu}\bar{\nu}} = \Omega_{\bar{\mu}\bar{\nu}} + \Omega_{\bar{\nu}\bar{\mu}}$, we obtain

$$i\frac{d}{dz}\hat{\phi}_H(\bar{\sigma}z) = \sum_{\bar{\mu}\bar{\nu}} \bar{\Omega}_{\bar{\mu}\bar{\nu}}(z) \alpha_{\bar{\sigma}\bar{\mu}} \hat{\phi}_H(\bar{\nu}z) + \sum_{\bar{\mu}} \int dx \lambda_{\bar{\mu}}(xz) \alpha_{\bar{\sigma}\bar{\mu}} \hat{\psi}_H^\dagger(xz) \hat{\psi}_H(xz).$$

(26)

Another equivalent form that will be used later can be obtained from Eq. (26) by using the idempotency relation $\delta_{\bar{\sigma}\bar{\sigma}'} = \sum_{\bar{\mu}} \alpha_{\bar{\sigma}\bar{\mu}} \alpha_{\bar{\mu}\bar{\sigma}'}$:

$$\sum_{\bar{\mu}} \left(i\alpha_{\bar{\sigma}\bar{\mu}}\frac{d}{dz} - \bar{\Omega}_{\bar{\sigma}\bar{\mu}}(z)\right) \hat{\phi}(\bar{\mu}z) = \int dx \lambda_{\bar{\sigma}}(xz) \hat{\psi}_H^\dagger(xz) \hat{\psi}_H(xz).$$

(27)

Equations (25) and (23) are independent on the initial state and play an important role in the development of the theory. As we explain in the subsequent sections, the whole structure of diagrammatic perturbation theory is encoded in them.

## 2.4    Many-Body Green's Functions

The ensemble averages of the correlators $\hat{G}_n$ and $\hat{\mathbb{D}}_n$ define corresponding $n$-particle Green's functions as

$$G_n(1, \ldots, n; 1', \ldots, n') \equiv \text{Tr}\left[\hat{\rho}\hat{G}_n(1, \ldots, n; 1', \ldots, n')\right]$$

(28)

$$\mathbb{D}_n(1, \ldots, 2n) \equiv \text{Tr}\left[\hat{\rho}\hat{\mathbb{D}}_n(1, \ldots, 2n)\right].$$

(29)

$n$-particle Green's functions directly provide ensemble averages of $n$-particle objects. With the time contour, as in Eq. (5), Green's functions can be written as

$$G_n(1, \ldots, n; 1', \ldots, n')$$
$$= (-i)^n \frac{\text{Tr}\left[\mathscr{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}(\bar{z})}\hat{\psi}(1)\ldots\hat{\psi}(n)\hat{\psi}^\dagger(n')\ldots\hat{\psi}^\dagger(1')\right\}\right]}{\text{Tr}\left[\mathscr{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}(\bar{z})}\right\}\right]}$$

(30)

$$\mathbb{D}_n(1, \ldots, 2n) = (-i)^n \frac{\text{Tr}\left[\mathscr{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}(\bar{z})}\hat{\phi}(1)\ldots\hat{\phi}(2n)\right\}\right]}{\text{Tr}\left[\mathscr{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}(\bar{z})}\right\}\right]}. \tag{31}$$

Note that the operators appearing in the time-ordering are Schrödinger operators. For future reference, we also define noninteracting Green's functions $g_n$ and $d_n$ by replacing $\hat{H}$ with $\hat{H}_0$. The form of Eqs. (30) and (31) is well-suited for perturbation expansions, which we will discuss in the next section.

The Green's functions $G_n$, $g_n$, $\mathbb{D}_n$, and $d_n$ fulfill the so-called Kubo-Martin-Schwinger (KMS) boundary conditions, which can be derived from Eqs. (30) and (31). For $G_n$, the conditions are $G(z_i, \ldots) = \pm G(z_f, \ldots)$ where $z_i/z_f$ is the initial/final time on the contour. The same holds for all arguments. Likewise, for $\mathbb{D}_n$: $\mathbb{D}_n(z_i, \ldots) = \mathbb{D}_n(z_f, \ldots)$, and similarly for $g_n$ and $d_n$.

The equations of motion for the Green's functions $G_n$ and $\mathbb{D}_n$ follow from the equations for their corresponding correlators, Eqs. (23) and (25).

$$\left[i\frac{\partial}{\partial z_k} - h(k)\right]G_n(1, \ldots, n; 1', \ldots, n')$$

$$= (-i)^n \text{Tr}\left[\hat{\rho}\mathscr{T}_c\left\{\hat{\psi}_H(1)\ldots\left[\hat{\psi}_H(k), \hat{H}_{\text{int}}(z_k)\right]_- \ldots\hat{\psi}_H(n)\hat{\psi}_H^\dagger(n')\ldots\hat{\psi}_H^\dagger(1')\right\}\right]$$

$$+ \sum_{j=1}^n (\pm)^{k+j}\delta(k; j')G_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, n; 1', \ldots, \overset{\sqcap}{j'}, \ldots, n'), \tag{32}$$

$$i\frac{\partial}{\partial z_k}\mathbb{D}_n(1, \ldots \bar{\sigma}z_k \ldots, 2n) - \sum_{\bar{\mu}\bar{\nu}}\alpha_{\bar{\sigma}\bar{\mu}}\bar{\Omega}_{\bar{\mu}\bar{\nu}}(z_k)\mathbb{D}_n(1, \ldots, \nu z_k, \ldots, 2n)$$

$$= (-i)^n \text{Tr}\left[\rho\mathscr{T}_c\left\{\hat{\phi}_H(1)\ldots\left[\hat{\phi}_H(k), \hat{H}_{\text{int}}(z_k)\right]_- \ldots\hat{\phi}_H(2n)\right\}\right] \tag{33}$$

$$+ \sum_{j=1, j\neq k}^{2n}\delta(z_k, z_j)\alpha_{\bar{\sigma}\bar{\mu}_j}\mathbb{D}_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, \overset{\sqcap}{j}, \ldots, 2n).$$

To further discuss these equations, we work out the commutators:

$$\left[\hat{\psi}_H(xz), \hat{H}_{\text{el-el}}(z)\right]_- = \int dx' v(x, x', z)\hat{\psi}_H^\dagger(x'z)\hat{\psi}_H(x'z)\hat{\psi}_H(xz)$$

$$\left[\hat{\psi}_H(xz), \hat{H}_{\text{el-bos}}(z)\right]_- = \sum_{\bar{\mu}}\lambda_{\bar{\mu}}(xz)\hat{\psi}_H(xz)\hat{\phi}_H(\bar{\mu}z)$$

$$\left[\hat{\phi}_H(\bar{\sigma}z), \hat{H}_{\text{el-bos}}\right]_- = \sum_{\bar{\mu}}\int dx \lambda_{\bar{\mu}}(xz)\alpha_{\bar{\sigma}\bar{\mu}}\hat{\psi}_H^\dagger(xz)\hat{\psi}_H(xz).$$

Let us start by considering only $\hat{H}_{\text{el-el}}$, that is, imagining that $\lambda = 0$. In this case, $\mathbb{D}_n = d_n$ and would only couple to $\mathbb{D}_{n-1} = d_{n-1}$. The electronic case is different; $G_n$ couples to $G_{n-1}$ but also to $G_{n+1}$.

In the case of a non-zero electron-electron and electron-boson interaction, $G_n$ would couple to $G_{n-1}$, and $G_{n+1}$, but also to mixed Green's functions consisting of a mix of $\hat{\psi}$, $\hat{\psi}^\dagger$, and $\hat{\phi}$. Likewise, $\mathbb{D}_n$ couples to $\mathbb{D}_{n-1}$ but also to mixed Green's functions. The equations of motion for these objects can also be derived in precisely the same way, but it will turn out that we will not need them for our purposes. For the derivation of mixed Green's functions, see Säkkinen (2016).

The full set of equations are referred to as the Martin-Schwinger hierarchy (Martin and Schwinger 1959). Together with the Kubo-Martin-Schwinger boundary conditions, we can uniquely obtain any $n$th-order correlation function. The exact solution is, of course, out of reach in most physical systems of relevance.

In the noninteracting case, the separate hierarchies of $g_n$ and $d_n$ will close on themselves: $g_n$ ($d_n$) couples only to $g_{n-1}$ ($d_{n-1}$). This allows for writing $g_n$ and $d_n$ in terms of $g_1$ and $d_1$, respectively. This is usually referred to as Wick's theorem and is the cornerstone of perturbation expansions. This will be further discussed below. First we will describe the general structure of perturbation theory for Green's functions.

## 2.5    Perturbation Expansion

Green's functions $G_n$ and $\mathbb{D}_n$ written using the time-contour ordering, Eqs. (30) and (31), permit an expansion in terms of $\hat{H}_{\text{int}}$. We focus on the single-particle $G \equiv G_1$, but the discussion for $\mathbb{D}_1$ or higher-order Green's functions is identical. Under the time-ordering operation $\hat{H}_0$, $\hat{H}_{\text{el-el}}$ and $\hat{H}_{\text{el-bos}}$ commute, and thus the exponentials of the three Hamiltonians can be separated. It then becomes natural to define the noninteracting average as $\text{Tr}\left[ \mathscr{T}_c \left\{ e^{-i \int_\gamma \mathrm{d}\bar{z} \hat{H}_0(\bar{z})} \cdots \right\} \right] = \langle \mathscr{T}_c \{ \cdots \} \rangle_0$ and expand $e^{-i \int_\gamma \mathrm{d}\bar{z} \hat{H}_{\text{el-el}}(\bar{z})} e^{-i \int_\gamma \mathrm{d}\bar{z} \hat{H}_{\text{el-bos}}(\bar{z})}$ in both numerator and denominator, resulting in

$$G(ab') = \frac{-i}{Z} \sum_{k,\tilde{\ell}=0}^{\infty} \frac{(-i)^{k+\tilde{\ell}}}{k!\tilde{\ell}!} \int_\gamma \mathrm{d}\bar{z}_1 \mathrm{d}\bar{z}_{\tilde{1}} \ldots \mathrm{d}\bar{z}_k \mathrm{d}\bar{z}_{\tilde{\ell}} \tag{34}$$

$$\langle \mathscr{T}_c \left\{ \hat{H}_{\text{el-el}}(\bar{z}_1) \ldots \hat{H}_{\text{el-el}}(\bar{z}_k) \hat{H}_{\text{el-bos}}(\bar{z}_{\tilde{1}}) \ldots \hat{H}_{\text{el-bos}}(\bar{z}_{\tilde{\ell}}) \hat{\psi}(a) \hat{\psi}^\dagger(b') \right\} \rangle_0,$$

where the partition function has a similar expansion

$$Z = \sum_{k,\tilde{\ell}=0}^{\infty} \frac{(-i)^{k+\tilde{\ell}}}{k!\tilde{\ell}!} \int_\gamma \langle \mathscr{T}_c \left\{ \hat{H}_{\text{el-el}}(\bar{z}_1) \ldots \hat{H}_{\text{el-el}}(\bar{z}_k) \hat{H}_{\text{el-bos}}(\bar{z}_{\tilde{1}}) \ldots \hat{H}_{\text{el-bos}}(\bar{z}_{\tilde{\ell}}) \right\} \rangle_0. \tag{35}$$

To distinguish between the fermion expansion and the boson one, we denote bosonic coordinates with tildes.

To evaluate Eq. (34), we need to be able to calculate time-ordered strings like $\mathrm{Tr}\left[\mathcal{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}_0(\bar{z})}\psi\ldots\psi\hat{\psi}^\dagger\ldots\hat{\psi}^\dagger\hat{\phi}\ldots\hat{\phi}\right\}\right]$ with an arbitrary number of operators. To make progress, it is crucial to note that the respective operators in $\hat{H}_0 = \hat{H}_{el} + \hat{H}_{bos}$ belong to different Hilbert spaces. As such, the eigenstates for $\hat{H}_0$ factorize into tensor products between the eigenstates of $\hat{H}_{el}$ and $\hat{H}_{bos}$. The noninteracting partition function $Z_0 = \mathrm{Tr}\left[\mathcal{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}_0(\bar{z})}\right\}\right]$ can thus be written as $Z_0 = \mathrm{Tr}_{el}\left[\mathcal{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}_{el}(\bar{z})}\right\}\right]\mathrm{Tr}_{bos}\left[\mathcal{T}_c\left\{e^{-i\int_\gamma d\bar{z}\hat{H}_{bos}(\bar{z})}\right\}\right] = Z_{el}Z_{bos}$. The same type of factorization allows us to simplify the expression above, as

$$\langle\mathcal{T}_c\left\{\psi\ldots\psi\hat{\psi}^\dagger\ldots\hat{\psi}^\dagger\hat{\phi}\ldots\hat{\phi}\right\}\rangle_0 = \langle\mathcal{T}_c\left\{\psi\ldots\psi\hat{\psi}^\dagger\ldots\hat{\psi}^\dagger\right\}\rangle_0\langle\mathcal{T}_c\left\{\hat{\phi}\ldots\hat{\phi}\right\}\rangle_0,$$
(36)

where it will be clear from the context if $\langle\mathcal{T}_c\{\cdots\}\rangle_0$ refers to the electronic or bosonic case.

The right-hand side of Eq. (36) is proportional to noninteracting Green's functions under suitable reorderings of the arguments. $2n$ operators yield $n$-particle Green's functions, $g_n$ for $\hat{\psi}$, $\hat{\psi}^\dagger$, and $d_n$ for $\hat{\phi}$. The expression for $G(ab')$, Eq. (34), with order $k$ for the electrons and $\tilde{\ell}$ for the bosons, yields the following time-ordered averages for the bosonic case (Eq. (31) with $\hat{H} \to \hat{H}_0$):

$$\langle\mathcal{T}_c\{\underbrace{\hat{\phi},\cdots,\hat{\phi}}_{\tilde{\ell}\text{ operators}}\}\rangle_0 = Z_{bos}i^{\tilde{\ell}/2}d_{\tilde{\ell}/2}(\tilde{1},\cdots,\tilde{\ell}).$$

For the fermionic operators, we have (Eq. (30) with $\hat{H} \to \hat{H}_0$)

$$\langle\mathcal{T}_c\{\underbrace{\hat{\psi}^\dagger\hat{\psi}^\dagger\hat{\psi}\hat{\psi},\cdots,}_{4k\text{ operators}}\underbrace{\hat{\psi}^\dagger\hat{\psi},\cdots,}_{2\tilde{\ell}\text{ operators}}\underbrace{\hat{\psi}(a)\hat{\psi}^\dagger(b')}_{2\text{ operators}}\}\rangle_0$$

$$= Z_{el}i^{2k+\tilde{\ell}+1}g_{2k+\tilde{\ell}+1}(a,\tilde{1},\cdots\tilde{\ell},1,1',\cdots;b',\tilde{1}^+,\cdots,\tilde{\ell}^+,1^+,1'^+,\cdots),$$

where we have been careful in reordering the operators, which results in an even number of permutations. To ease notation, we defined $g(\tilde{1},\cdots) = g(x_{\tilde{1}},z_{\tilde{1}},\cdots)$.

From these considerations, we can rewrite Eq. (34) as

$$G(ab') = \frac{Z_0}{Z}\sum_{k,\tilde{\ell}=0}^\infty \frac{i^k}{2^k k!}\frac{i^{\tilde{\ell}/2}}{\tilde{\ell}!}\int v(11')\cdots v(kk')\int \lambda(\tilde{1})\cdots\lambda(\tilde{\ell})$$

$$d_{\tilde{\ell}/2}(\tilde{1},\ldots,\tilde{\ell})g_{2k+\tilde{\ell}+1}(a,\tilde{1},\ldots,\tilde{\ell},1,1',\ldots;b',\tilde{1}^+,\ldots,\tilde{\ell},1^+,1'^+,\ldots),$$
(37)

where we defined $v(11') = \delta(z, z')v(x, x', z)$. The first type of integral is over $1, 1', \ldots, k, k'$, and the second type of integral is over $\tilde{1}, \ldots, \tilde{\ell}$, by which we mean $\int_\gamma d\bar{z}_{\tilde{1}} \cdots d\bar{z}_{\tilde{\ell}} \int dx_{\tilde{1}} \cdots dx_{\tilde{\ell}} \sum_{\bar{\mu}_1, \cdots, \bar{\mu}_{\tilde{\ell}}}$. To ease the notation, we write $\lambda(\tilde{\ell}) = \lambda_{\bar{\mu}_{\tilde{\ell}}}(x_{\tilde{\ell}}, z_{\tilde{\ell}})$ and $d(\tilde{1}, \cdots) = d(\bar{\mu}_{\tilde{1}}, z_{\tilde{1}}, \cdots)$. The factor $1/2^k$ comes from the factor $1/2$ in $\hat{H}_{\text{el-el}}$. The factor $i^1$ conveniently canceled the $(-i)$ factor in the definition of $G$. A similar equation holds for $\mathbb{D} = \mathbb{D}_1$ and all higher-order Green's functions $G_n$ and $\mathbb{D}_n$.

As such, perturbation theory for mixed fermion-boson systems is reduced to finding practical expressions for the separate non-interacting electronic and bosonic $n$-particle Green's functions, $g_n$ and $d_n$. The expressions are provided by Wick's theorem, which is the topic of the next section.

## 2.6    Wick's Theorem

Wick's theorem provides a recipe of how to write $g_n$ as products of $g_1$:s and $d_n$ as products of $d_1$. The original theorem (Wick 1950) was an operator identity which related time-ordered products to normal-ordered ones and contractions. Although being the cornerstone of zero-temperature formalism, where normal-ordered products vanish, the theorem is not immediately applicable for finite temperature or systems out of equilibrium. Also, Wick's discussion was focused on creation and destruction operators.

Here we will show how a Wick-like decoupling works for the general case for $g_n$ and $d_n$. All cases (zero temperature, finite temperature in equilibrium, non-equilibrium) follow from using the general time contour and then deforming it (Stefanucci and van Leeuwen 2013). The decoupling follows from the Martin-Schwinger hierarchy, Eqs. (32) and (33), with $\hat{H}_{\text{int}} = 0$. The decoupling is different for $g_n$ and $d_n$, essentially since $g_n$ contains two types of non-Hermitian operators ($\hat{\psi}$ and $\hat{\psi}^\dagger$), while $d_n$ contains only one type of Hermitian operator ($\hat{\phi}$). As such, we discuss them separately, starting with the more familiar case of $g_n$.

### 2.6.1    Wick's Theorem for $g_n$

The determining relation for $g_n$ is the Martin-Schwinger hierarchy for $G_n$, Eq. (32), with $\hat{H}_{\text{int}} = 0$, and the KMS boundary conditions:

$$\left[ i\frac{\partial}{\partial z_k} - h(1) \right] g_n = \sum_{j=1}^{n} (\pm)^{k+j} \delta(k; j') g_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, n; 1', \ldots, \overset{\sqcap}{j'}, \ldots, n').$$

(38)

The Martin-Schwinger hierarchy for $g_n$ couples only to $g_{n-1}$, and thus by iteration, we can find the solution to Eq. (38) in terms of single-particle Green's functions $g(1; 1') \equiv g_1(1; 1')$. The explicit solution is given by a permanent/determinant for bosons/fermions of $g(1; 1')$ (Martin and Schwinger 1959; Bruus and Flensberg 2004; Stefanucci and van Leeuwen 2013):

$$g_n(1, \ldots, n; 1', \ldots, n') = \begin{vmatrix} g(1; 1') & \ldots & g(1; n') \\ \vdots & & \ldots \\ g(n; 1') & \ldots & g(n; n') \end{vmatrix}_{\pm}, \tag{39}$$

where $g(1; 1')$ satisfies (Eq. (38) with $n = 1$)

$$\left( i \frac{\partial}{\partial z_1} - h(1) \right) g(1; 1') = \delta(1; 1'). \tag{40}$$

The permanent/determinant Eq. (39) satisfies the Martin-Schwinger equations for $g_n$, Eq. (38), as can be checked by applying the differential operator $\left[ i \frac{\partial}{\partial z_k} - h(1) \right]$. Furthermore, the KMS boundary conditions are satisfied since $g$ satisfies them.

As an example, we consider the decoupling for $g_2$:

$$g_2(1, 2; 1', 2') = g(1; 1')g(2; 2') \pm g(1; 2')g(2; 1') \tag{41}$$

which is depicted in Fig. 2. The correct sign is taken into account by simply counting the number of times lines cross each other.

We can also write the decoupling recursively. For $g_3$, expanded along 1

$$g_3(1, 2, 3; 1', 2', 3') = g(1; 1')g_2(2, 3; 2', 3') \pm g(1; 2')g_2(2, 3; 1', 3')$$
$$+ g(1; 3')g_2(2, 3; 1', 2'), \tag{42}$$

which is depicted in Fig. 3. In the recursive form along $k$, it is immediately clear that $g_3$ satisfies the Martin-Schwinger hierarchy by applying $\left[ i \frac{\partial}{\partial z_k} - h(1) \right]$.



**Fig. 2** Wick's theorem for noninteracting two-particle Green's function $g_2$, Eq. (41)
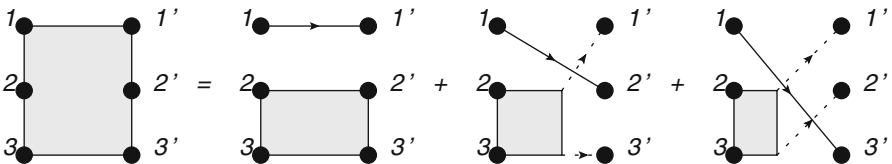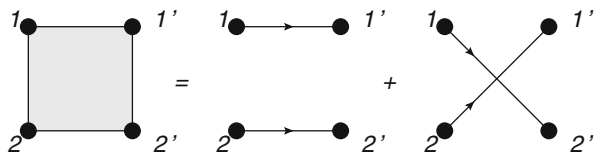


**Fig. 3** Wick's theorem for noninteracting three-particle Green's function $g_3$, Eq. (42). The second diagram on the right-hand side gives a contribution $(\pm 1)$, having one crossing, while the right diagram gives $(\pm 1)^2 = 1$. We have prolonged the edges of the $g_2$:s with a thin dashed line to the correct vertices since we can then easily see which sign the diagram has by counting the number of crossings

### 2.6.2 Wick's Theorem for $d_n$

The determining relation for $d_n$ is the Martin-Schwinger hierarchy for $\mathbb{D}_n$, Eq. (33), with $\hat{H}_{\text{int}} = 0$, and the KMS boundary conditions:

$$\sum_{\bar{\mu}} \left[ i\alpha_{\bar{\sigma}\bar{\mu}} \frac{\partial}{\partial z_k} - \bar{\Omega}_{\bar{\sigma}\bar{\mu}}(z_k) \right] d_n(1, \ldots, \bar{\mu}z_k, \ldots, 2n)$$

$$= \sum_{j=1, j\neq k}^{2n} \delta(z_k, z_j)\delta_{\bar{\sigma}\bar{\mu}_j} d_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, \overset{\sqcap}{j}, \ldots, 2n). \tag{43}$$

The hierarchy for $d_n$, Eq. (43), clearly resembles the hierarchy for the electronic case, $g_n$, Eq. (38). The structure becomes more clear by defining the operator in front of $d_n$ on the first line as an (discrete) integral operator $L_k$, leading to

$$L_k d_n(1, \ldots, k, \ldots, 2n) = \sum_{j=1, j\neq k}^{2n} \delta(k, j) d_{n-1}(1, \ldots, \overset{\sqcap}{k}, \ldots, \overset{\sqcap}{j}, \ldots, 2n). \tag{44}$$

This is the formal analogue to Eq. (38), and as such it is not surprising that Wick's theorem for $g_n$ and $d_n$ is very similar. The difference is that for $d_n$, all operators need to be contracted with each other, as there is no separation into cases as it was for creation and destruction operators.

Another feature that is different from the electronic case is the appearance of an odd number of fields, when $n$ is a half-integer. From Eq. (44), we see that integer $n$ connects with integer $n - 1$ and half-integer $n$ connects with half-integers $n - 1$. As such, we have two separate hierarchies of equations. Furthermore, we have that $d_n = 0$ when $n$ is a half-integer. This can be seen by considering the ensemble average of the equation of motion for $\hat{\phi}_H$, Eq. (27), in the interacting case:

$$\sum_{\bar{\mu}} \left( i\alpha_{\bar{\sigma}\bar{\mu}} \frac{\mathrm{d}}{\mathrm{d}z} - \bar{\Omega}_{\bar{\sigma}\bar{\mu}}(z) \right) \phi(\bar{\mu}z) = \int dx \lambda_{\bar{\sigma}}(xz)G(x, z, x, z^+). \tag{45}$$

This equation can be solved, and the solution is given by

$$\phi(\bar{\sigma}, z) = -i \sum_{\bar{\mu}} \int dx \int_{\gamma} d\bar{z} \, d(\bar{\sigma}z; \bar{\mu}\bar{z})\lambda_{\bar{\mu}}(x, \bar{z})G(x, \bar{z}, x, \bar{z}^+), \tag{46}$$

where $d = d_1$ is the bosonic Green's function of Eq. (44) with $n = 1$ that fulfills

$$\sum_{\bar{\mu}} \left[ i\alpha_{\bar{\sigma}\bar{\mu}} \frac{\partial}{\partial z_1} - \bar{\Omega}_{\bar{\sigma}\bar{\mu}}(z_1) \right] d(\bar{\mu}z_1, 2) = \delta(z_1, z_2)\delta_{\bar{\sigma}\bar{\mu}_2}. \tag{47}$$

In the non-interacting state, we have immediately that $\langle \phi(\bar{\sigma}, z) \rangle_0 = 0$, that is, we have $d_{1/2} = 0$. Moreover, the Martin-Schwinger hierarchy couples half-integers to half-integers. Consider $n = 3/2$:

$$L_1 d_{3/2}(1, 2, 3) = \delta(1, 2)d_{1/2}(3) + \delta(1, 3)d_{1/2}(2) = 0$$

We have that $d_{3/2} = 0$ is a solution, which fulfill the KMS boundary conditions. As such, this is the only solution. By induction, $d_n = 0$ for half-integers. As such, from now on, we only consider $n$ being an integer in the noninteracting case.

As in the case of $g_n$, the equation of motion for $d_n$ only connects to lower-order Green's functions, making it plausible that we can write the solution to Eq. (43) in terms of single-particle Green's functions $d(1, 2) \equiv d_1(1, 2)$. For $d_n$; the solution to Eq. (43) is given by the so-called Haffnian (Säkkinen 2016):

$$d_n(1 \ldots 2n) = d(12)d(34) \cdots d(2n - 1, 2n) + d(13)d(24) \cdots d(2n - 1, 2n) + \ldots,$$
(48)

where ... denote the rest of the permutations of 1 to the set $\{4, \ldots, 2n\}$.

As examples of the decoupling of $d_n$, $d_2$ is given by

$$d_2(1234) = d(12)d(34) + d(13)d(24) + d(14)d(23).$$
(49)

The solution can also be written in a recursive form, which can be seen as expanding the Haffnian around a point. For $d_3$, expanded around the point 1

$$d_3(123456) = d(12)d_2(3456) + d(13)d_2(2456) + d(14)d_2(2356)$$
$$+ d(15)d_2(2346) + d(16)d_2(2345).$$
(50)

The recursive form of Eq. (50) makes it clear that $d_3$ satisfies Eq. (43) with $k = 1$. The expansion for $d_3$ is shown in Fig. 4, which exemplifies that the Haffnian can be visualized in terms of the so-called chord diagrams. We also note that the Haffnian is commonly written in a more symmetric form, due to the symmetry $d(12) = d(21)$:

$$d_n(1 \ldots 2n) = \frac{1}{2^n n!} \sum_P \prod_{i=1}^{n} d(P(2i - 1), P(2i)),$$
(51)

where $P$ denotes permutations of $\{1, \ldots 2n\}$.

We thus again note the crucial difference between Wick's theorem for $g_n$ and $d_n$: For $g_n$ we need to connect unprimed coordinates to primed coordinates in all possible ways. However, for $d_n$, there is no distinction between primed and unprimed coordinates, and we need to connect all coordinates in all possible ways. It is thus natural that for $g_n$ we organize the vertices in two columns, while for $d_n$ it is more natural to put the vertices on a circle. It is important to note, however, that the decoupling, Wick's theorem, is in both cases the solution to the corresponding non-interacting Martin-Schwinger hierarchy with the proper Kubo-Martin-Schwinger
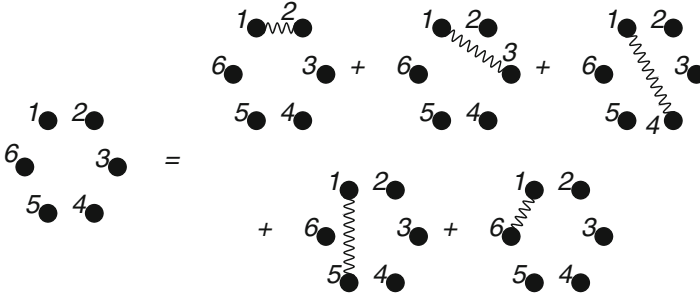
**Fig. 4** Pictorial description of Wick's theorem for $d_3$, Eq. (50), when expanding around point 1. Dots connected with a wiggly line denote $d_1$, while non-connected dots denote higher-order Green's function. Note that we do not need to take into account the number of possible crossings, since the particles here are bosons

boundary conditions. As stated before, the time contour construction permits a finite-temperature, non-equilibrium perturbation theory. By specializing to different contours, we obtain the zero-temperature formalism by an adiabatic switch of the interaction or the Matsubara formalism by considering the vertical track only. This shows the elegance of the contour formalism for mixed electron-boson systems.

## 3    Diagrammatic Perturbation Theory

Wick's theorems provided explicit expressions for noninteracting Green's functions $g_n$ and $d_n$, which were the building blocks in the perturbation expansion for many-body Green's functions $G_n$ and $\mathbb{D}_n$, via equations like Eq. (37).

Equipped with Wick's theorem, we can write the expansion for $G(ab')$, Eq. (37), in diagrams. It turns out, however, that many of the resulting diagrams are topologically equivalent and as such we only need to sum over the topologically inequivalent diagrams. This will cancel the combinatorial factors in Eq. (37). Furthermore, we only need to sum over connected diagrams, meaning that the internal vertices need to connect to the outer vertices (those vertices not integrated over). The disconnected diagrams will cancel against the expansion for $Z$ in Eq. (35). For a more in-depth discussion, see, for example, Säkkinen (2016). The resulting formulas for $G$ and $D$ become

$$G(ab') = \sum_{k,\tilde{\ell}=0}^{\infty} i^{k+\tilde{\ell}} \int v(11') \cdots v(kk') \int \lambda(\tilde{1}) \cdots \lambda(2\tilde{\ell})$$

$$d_{\bar{1}}(\tilde{1}, \ldots, 2\tilde{\ell}) g_{2k+2\tilde{\ell}+1}(a, \tilde{1}, \cdots, \tilde{\ell}, 1, 1', \ldots; b', \tilde{1}, \cdots, \tilde{\ell}, 1^+, 1'^+, \ldots)|_{\text{t.in.c}}$$

$$(52)$$
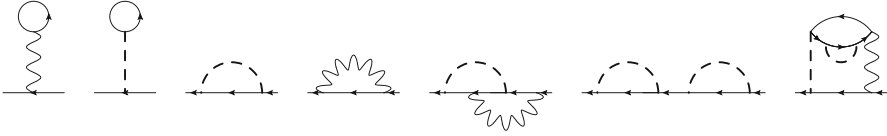
**Fig. 5** Example diagrams of $G(ab')$. Wavy lines denote $d$ propagators, while dashed lines denote the pair interaction $v(x, x', z)$. Solid straight lines denote $g$ propagators

$$
\mathbb{D}(ab) = \sum_{k,\tilde{\ell}=0}^{\infty} i^{k+\tilde{\ell}} \int v(11') \cdots v(kk') \int \lambda(\tilde{1}) \cdots \lambda(2\tilde{\ell})
$$

$$
d_{\tilde{\ell}+1}(ab\tilde{1}\ldots 2\tilde{\ell})g_{2k+2\tilde{\ell}}(\tilde{1}, \cdots, \tilde{\ell}, 1, 1', \ldots; \tilde{1}, \cdots, \tilde{\ell}, 1^+, 1'^+, \ldots)|_{\text{t.in.c}}, \tag{53}
$$

where the abbreviation (t.in.c) denotes that we consider only the topologically inequivalent and connected diagrams when expanding $g_n$ and $d_n$ in permanents/determinants and Haffnians. The same considerations can be applied to higher-order Green's functions without any formal complication. As such, perturbation theory for coupled electron-boson systems is now completely well-defined.

## 3.1    Skeleton Diagrams and the Kadanoff-Baym Equations

The Feynman rules for the fermionic and bosonic $n$-particle Green's functions completely determine these functions in terms of the propagators $g$ and $d$. We can, however, reduce the number of distinct diagrammatic terms considerably by expressing them in terms of the so-called dressed propagators. Let us start by discussing the dressing of the fermion propagator $G$.

We first give some examples of the diagrammatic expansion of $G(ab')$, in Fig. 5. We note that many diagrams will contain repetitions of smaller blocks. This motivates us to define the concept of an irreducible self-energy $\Sigma$, which consists of the set of diagrams contributing to $G$ that, after the removal of the ingoing and outgoing $g$ lines, cannot be cut into two pieces by removal of a $g$ line. With these definitions we can write $G$ as a repetition in series of irreducible pieces:

$$
G = g + g\Sigma g + g\Sigma g\Sigma g + \ldots = g + g\Sigma G \tag{54}
$$

in which the products in these expressions consist of repeated convolutions, i.e., $ab(1, 2) = \int_\gamma d3\, a(1, 3)b(3, 2)$. The self-energy $\Sigma[g, d]$ is itself expressed as an infinite sum of irreducible diagrams in terms of $g$ and $d$. Examples of diagrams for $\Sigma[g, d]$ are shown in Fig. 6.

Within these $\Sigma$ diagrams themselves, we can also distinguish self-energy insertions (e.g., the last diagram in Fig. 6), i.e., pieces of diagrams that can be made
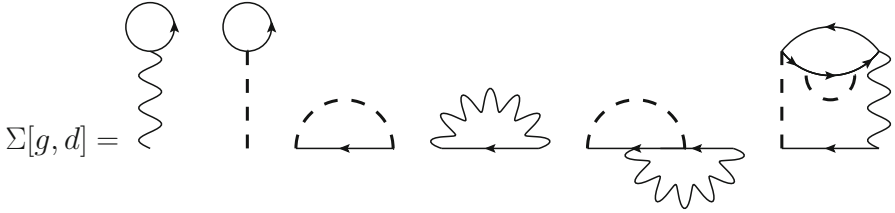
**Fig. 6** Example diagrams of the electronic self-energy $\Sigma[g, d]$. The last diagram contains a self-energy insertion
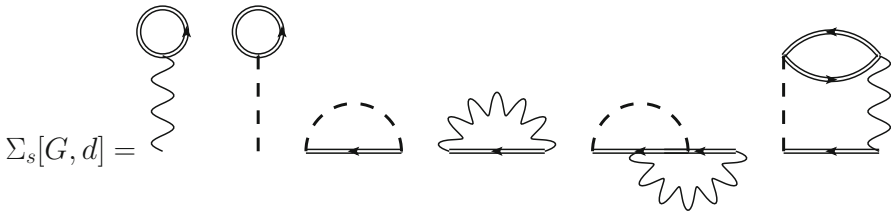


**Fig. 7** Example diagrams of the skeletonic electronic self-energy $\Sigma_s[G, d]$

disjoint of the rest of the diagram by cutting two $g$ lines. The diagrams obtained by the removal of all self-energy insertions are the so-called skeleton diagrams. The full set of $\Sigma$ diagrams can be obtained by dressing the skeleton graphs with self-energy insertions in all possible ways. This amounts to the replacement of the bare $g$ lines by the full propagators $G$ in every diagram. We can therefore write $\Sigma[g, d] = \Sigma_s[G, d]$ where we used the subscript $s$ to denote that $\Sigma_s$ is the sums of all skeleton diagrams in terms of $G$ and $d$. We can therefore write

$$G = g + g\,\Sigma_s[G, d]G. \tag{55}$$

Examples of diagrams for $\Sigma_s[G, d]$ are shown in Fig. 7.

Let us now proceed with the discussion of the bosonic Green's function $\mathbb{D}(ab)$. The diagrammatic structure comes from expanding $g_n$ and $d_n$ in Eq. (53), and example diagrams are shown in Fig. 8.

New types of diagrams, the double "tadpole" diagrams (top middle in Fig. 8), occur for $\mathbb{D}(ab)$ that were not present for $G(ab)$, and as such we discuss them first. The tadpoles are in fact the diagrammatic expansion of the half-body term $\phi(a)$. All the tadpole diagrams can be conveniently summed up, which is done in Fig. 9. The diagrammatic rules yield

$$\phi(\bar{\sigma}, , z) = -i \sum_{\bar{\mu}} \int dx \int_\gamma d\bar{z}\, d(\bar{\sigma}z; \bar{\mu}\bar{z})\lambda_{\bar{\mu}}(x, \bar{z})G(x, \bar{z}, x, \bar{z}^+), \tag{56}$$
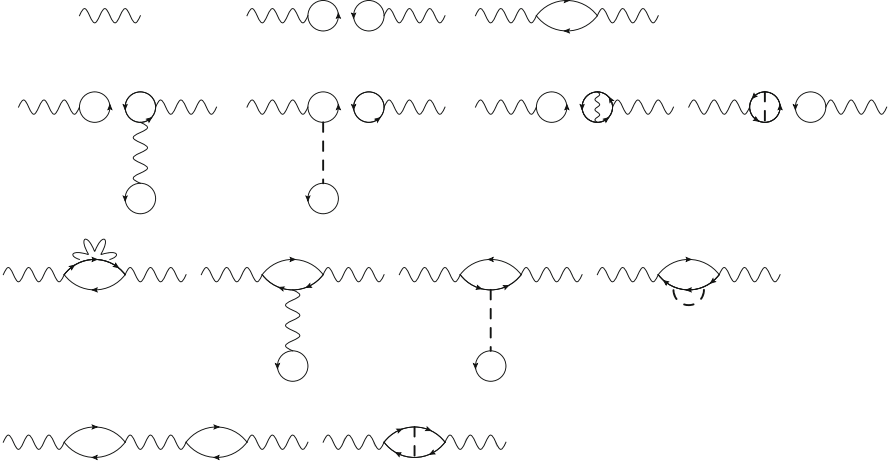
**Fig. 8** Example diagrams of $\mathbb{D}(ab)$. The top middle diagram is referred to as the double tadpole diagram
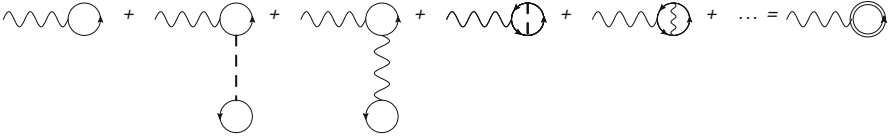


**Fig. 9** Summation of all tadpole diagrams, which yields $\phi(a)$, Eq. (56)
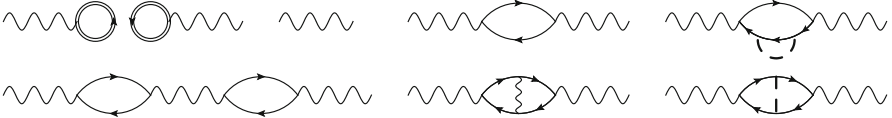


**Fig. 10** Example diagrams of $\mathbb{D}(ab)$ after resumming the double tadpole diagrams

a result that we already obtained from the equations of motion for $\phi(a)$, Eq. (46). As such, we can write the expansion for $\mathbb{D}$ as

$$\mathbb{D}(1, 2) = \phi(1)\phi(2) + d(1, 2) + (d\Pi d)(1, 2) + (d\Pi d\Pi d)(1, 2) + \dots, \quad (57)$$

where we also introduced a bosonic self-energy $\Pi[g, d]$ as the collection of diagrams which cannot be made disjoint by cutting two $d$ lines. Some example diagrams are shown in Fig. 10.

Differently from the electronic case, $\mathbb{D}(12)$ does NOT fulfill a Dyson equation, due to the presence of $\phi(a)\phi(b)$. As such, we are motivated to define a new type of Green's function $D(ab) = \mathbb{D}(ab) - \phi(a)\phi(b)$ which has these terms subtracted. $D(ab)$ can be defined using the so-called fluctuation operators,

$$\Pi_s[G, d] =$$ 
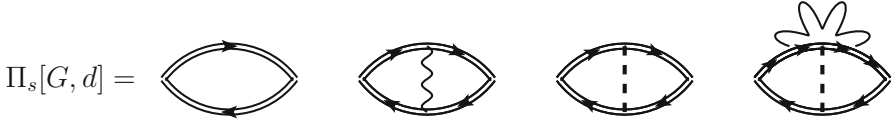
**Fig. 11** Example diagrams of the bosonic skeletonic self-energy $\Pi_s[G, d]$
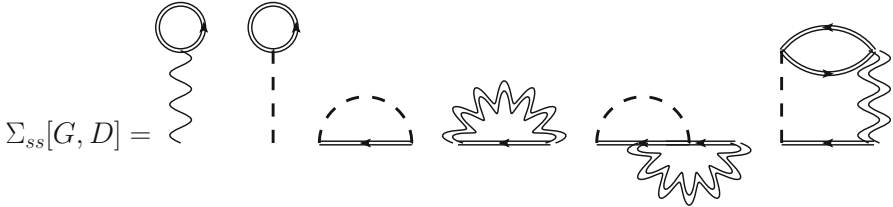
$$\Sigma_{ss}[G, D] =$$ 

**Fig. 12** Example diagrams of the double skeletonic electronic self-energy $\Sigma_{ss}[G, D] = \Sigma_H[G] + \Sigma_{xc}[G, D]$, see Eq. (63). The first two diagrams, the tadpole diagrams, make up the time-local $\Sigma_H$

$\Delta\hat{\phi}_H(a) = \hat{\phi}_H(a) - \phi(a)$, with $\langle\Delta\hat{\phi}_H(a)\rangle = 0$. We then have that $D(ab) = -i\text{Tr}\left[\mathcal{T}_c\left\{\hat{\rho}\Delta\hat{\phi}_H(a)\Delta\hat{\phi}_H(b)\right\}\right]$. The Fluctuation Green's function $D(ab)$ fulfills a Dyson equation, obtained by rewriting Eq. (57) as

$$D = d + d\Pi d + d\Pi d\Pi d + \ldots = d + d\Pi[g, d]D. \tag{58}$$

We can now proceed as above and remove all fermionic self-energy insertions inside $\Pi[g, d]$ diagrams and form skeleton diagrams and replace the bare $g$ by the dressed $G$ as $\Pi_s[G, d] = \Pi[g, d]$. This is shown in Fig. 11.

We note now that we can continue this way and also remove all bosonic self-energy insertions inside the $\Pi_s$ diagrams and dress these diagrams with the full $D$ line such that $\Pi_{ss}[G, D] = \Pi_s[G, d]$ where $\Pi_{ss}$ are the self-energy diagrams that are skeletonic with respect to removal of both fermionic and bosonic self-energy insertions. A similar procedure can be carried out for the fermionic self-energy, where we can write $\Sigma_{ss}[G, D] = \Sigma_s[G, d]$, except that we must not dress the time-local diagrams, as shown in Fig. 12. In conclusion, we obtain the following set of equations

$$G = g + g\Sigma_{ss}[G, D]G \tag{59}$$

$$D = d + d\Pi_{ss}[G, D]D. \tag{60}$$

Example diagrams are shown in Fig. 12 for $\Sigma_{ss}[G, D]$ and Fig. 13 for $\Pi_{ss}[G, D]$.

The Dyson equations, Eqs. (59) and (60), can be rewritten in a form often used in time-dependent calculations, by using the equations of motion for bare Green's functions:
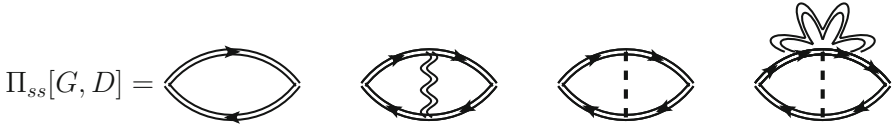
$$\Pi_{ss}[G, D] = \text{}$$

**Fig. 13** Example diagrams of the double skeletonic bosonic self-energy $\Pi_{ss}[G, D]$

$$\left[i\partial_{z_1} - h(1)\right] G(1, 2) = \delta(1, 2) + \int_\gamma d3 \, \Sigma_{ss}[G, D](1, 3)G(3, 2) \tag{61}$$

$$\left[i\alpha\partial_{z_1} - \bar{\Omega}(1)\right] D(1, 2) = \delta(1, 2) + \int_\gamma d3 \, \Pi_{ss}[G, D](1, 3)D(3, 2). \tag{62}$$

These are self-consistent equations for the fermion and boson propagator, once an approximation for $\Sigma_{ss}$ and $\Pi_{ss}$ in terms of skeleton graphs has been chosen. Two similar adjoint equations involving the time derivative with respect to $z_2$ can be derived in a similar way. These equation are known as the Kadanoff-Baym equations (KBE), together with the KMS boundary conditions discussed earlier.

The solution of the KBE yields a wealth of information. Apart from non-equilibrium spectral functions and the total energy, the expectation value of any one-body operator can be obtained from the solution. Since external time-dependent fields are treated explicitly, the approach allows, for example, for the calculation of higher response functions. For the case of linear response functions, this provides an alternative for the Bethe-Salpeter equation while at the same time admitting sophisticated memory kernels which are difficult to treat in the standard approach.

The solution of the KBE requires full self-consistency, as the self-energies depend on Green's functions that are themselves solutions of the equations. This feature can be a numerical disadvantage, but it also has two important theoretical advantages. First of all, it removes the dependence of the result on bare Green's functions, allowing for an unambiguous calculation of observables. Secondly, by a proper diagrammatic choice of the self-energies, we can make sure that the observables calculated from the Green's functions satisfy the conservation laws for energy, momentum, and particle number. For this purpose, the self-energies must be chosen in a so-called $\Phi$-derivable manner. To define this properly, it is convenient to split off the spatially and temporally local Hartree part $\Sigma_H$ of the fermionic self-energy and to define the remaining exchange-correlation part $\Sigma_{xc}$ by

$$\Sigma_{ss}(1, 2) = \Sigma_H(1, 2) + \Sigma_{xc}(1, 2). \tag{63}$$

The Hartree part is given by the two tadpole diagrams of Fig. 12. Now we define a functional $\Phi[G, D]$ by closing each skeleton diagram for $\Sigma_{xc}$ with a $G$ line which produces a set of vacuum diagrams. The procedure is explained in more detail in Fig. 14.

$$\Phi[G, D] = \frac{1}{2} \bigcirc + \frac{1}{2} \bigcirc + \frac{1}{4} \bigcirc + \frac{1}{2} \bigcirc + \frac{1}{4} \bigcirc + \frac{1}{4} \bigcirc \bigcirc + \frac{1}{2} \bigcirc \bigcirc$$
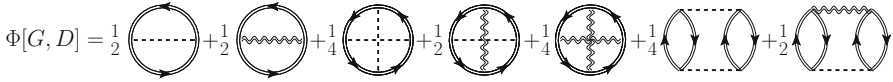
**Fig. 14** Example diagrams of $\Phi[G, D]$ yielding $\Sigma_{ss}[G, D]$ and $\Pi_{ss}[G, D]$. As examples, the operation $\delta/\delta G$, i.e., removing a $G$ line, applied to the first two diagrams gives the third and fourth diagram for $\Sigma_{ss}[G, D]$ in Fig. 12 (Fock-like, or exchange-like, diagrams). The operation $-2\delta/\delta D|_S$ applied to the second diagram yields the first diagram for $\Pi_{ss}[G, D]$ in Fig. 13

Each of the topologically different vacuum diagrams will also be multiplied with a symmetry factor $1/N_S$ where $N_S$ is the number of equivalent $G$ lines that yield the same self-energy diagram by their respective removal (the relevant fact that the different classes of equivalent lines in a vacuum diagram are equal in size can be proven as in Stefanucci and van Leeuwen 2013). We furthermore introduce an additional minus sign since the removal of a $G$ line from a vacuum diagram changes the number of fermion loops by one. By construction the $\Phi$-functional has the property that $\Sigma_{xc}(1, 2) = \delta\Phi/\delta G(2, 1)$. Along the lines of Stefanucci and van Leeuwen (2013), we can further prove that $\Pi_{ss}(1, 2) = -2\delta\Phi/\delta D(1, 2)|_S$ where the subscript $S$ refers to the symmetrized derivative $[\delta/\delta D(12) + \delta/\delta D(21)]/2$.

So far we dealt with the exact self-energies, but in practice we have to consider approximate self-energies. We say that such self-energies $\Sigma_{xc}$ and $\Pi_{ss}$ are $\Phi$-derivable, whenever there exists a functional $\Phi[G, D]$ such that

$$\Sigma_{xc}(1, 2) = \frac{\delta\Phi}{\delta G(2, 1)} \tag{64}$$

$$\Pi_{ss}(1, 2) = -2\frac{\delta\Phi}{\delta D(1, 2)}|_S. \tag{65}$$

In practice the corresponding $\Phi$-functional is obtained by selection of an appropriate set of sub-diagrams of the exact $\Phi$-functional.

The significance of this procedure lies in the fact that the $\Phi$-functional has invariance properties with regard to certain space-time and gauge transformations that imply the satisfaction of corresponding conservation laws for Green's functions obtained from the KBE in which the $\Phi$-derivable self-energies are used (Baym and Kadanoff 1961; Baym 1962). We therefore have a way to automatically guarantee the satisfaction of conservation laws for approximate self-energies. We finally mention that it is also possible to develop partially self-consistent approaches in which, for example, only number conservation is guaranteed (Karlsson and van Leeuwen 2016). Such approaches are of interest when full self-consistency is hard to achieve, for example, due to computational limitations.

We finally note that for the purely electronic case, it is very common to dress also the two-body interaction $v(x, y, z)$, to obtain expansions in terms of the so-called screened interaction $W$. This construction, to lowest order, constitutes the often-used $GW$ approximation (Hedin 1965). The redressed $\Phi$-diagrams are referred to as the $\Psi[G, W]$-functional (Almbladh et al. 1999). In the presence of phonons, similar types of resummations can be done (Hedin and Lundqvist 1970).

We should mention that the KBE for practical applications are typically, by selecting time variables on various branches of the contour, rewritten in terms of real-time, imaginary time, and mixed real-imaginary time components using the so-called Langreth rules. This is more a technical issue, and we refer to the literature (Dahlen and van Leeuwen 2007; Stefanucci and van Leeuwen 2013) for details. In practice these equations are first solved on the vertical imaginary track of the contour, after which a time propagation in the double real-time and real-imaginary time variables can be performed. In practice the KBE are typically converted, using a suitable basis, into time-dependent matrix equations which can be solved by employing suitable time-stepping techniques in the double-time plane (Stan et al. 2009).

## 4    Conclusions

The use of non-equilibrium Green's function methods has seen an explosive rise in the last two decades, mainly due to developments in nano-science and nonlinear pump-probe laser technologies that allow for in-depth studies of non-equilibrium systems.

One of the large research fields that has been developed over the last years is the study of molecular transport which studies charge transport through single molecules and nano-junctions (Stefanucci and Almbladh 2004; Datta 2005; Cuevas and Scheer 2010) attached to metallic leads. The small size of the system makes it necessary to have a detailed description of the electronic structure in which a correct description of the Coulomb interactions between the electrons in the system is crucial. The Kadanoff-Baym approach is well-suited to study time-dependent quantum transport systems. If the macroscopic leads are treated as effectively non-interacting, the leads can be incorporated non-perturbatively into the KBE using the so-called embedding self-energy, which is simply added to the many-body self-energy. The first applications of KBE to transient phenomena in correlated charge transport of electrons are by now a decade old (Myöhänen et al. 2008, 2009; Puig von Friesen et al. 2010; Uimonen et al. 2011) with later developments of numerically efficient but simplified approaches (Latini et al. 2014) based on the so-called generalized Kadanoff-Baym ansatz (Lipavský et al. 1986). These approaches have given new insights into various non-equilibrium phenomena such as transients and memory effects (Myöhänen et al. 2008), as well as bistability (Uimonen et al. 2010; Khosravi et al. 2012), image charge effects (Myöhänen et al. 2012), bias-induced gap closings (Myöhänen et al. 2009), and other correlation effects in the non-equilibrium regime.

Another research field that developed quickly is that of strong-field pump-probe laser spectroscopy of single molecules in strong laser fields. In these systems, there is a considerable ionization yield, and the use of non-equilibrium Green's functions is very well-suited to describe these systems using similar embedding techniques as in the quantum transport case, in which the embedding is now done for the ionization continuum rather than the macroscopic leads. This allows for treating strong external fields non-perturbatively (Perfetto et al. 2015). The formalism has been applied to the first-principle study of ultrafast phenomena in large organic

molecules such as the early-stage density oscillation of the electronic charge in the photoexcited phenylalanine amino acid (Perfetto et al. 2018) as well as the electronic charge separation in photoexcited donor-$C_{60}$ complexes (Boström et al. 2018).

All these systems were purely electronic. However, recent years have seen the development of the KBE for coupled electron-boson systems (Säkkinen et al. 2015a, b) as well as the study of the KBE (Tuovinen et al. 2016) for heat transport (Arrachea et al. 2012; Wang et al. 2014). In these systems we not only deal with electrons but also with various bosonic systems. The KBE for electron-boson systems have been used to study polaron formation in coupled electron-phonon systems (Säkkinen et al. 2015a, b), spin transport for electron-magnon systems (Mahfouzi and Nikolić 2014), and pump-probe spectroscopies in electron-plasmon systems (Schüler et al. 2016), and there is a considerable activity toward the study of electron-photon systems in cavity quantum electrodynamics (Ruggenthaler et al. 2014; Flick et al. 2017) with various connections to time-dependent density functional theory. This research field is growing rapidly, and the KBE method and its simplifications are becoming standard tools; there is a transition of applications from model systems to realistic systems (Perfetto et al. 2015; de Melo and Marini 2016; Schlünzen et al. 2016).

Pump-probe and nonlinear spectroscopy on molecules and solids is also experimentally a growing research field, and therefore there is an increasing need for theoretical methods to interpret and describe new experiments. In this work, we outlined a suitable formalism to achieve this goal. The formalism has already been used successfully in the lowest orders, and many promising new applications and further developments of the theory are within reach.

# References

Almbladh CO, von Barth U, van Leeuwen R (1999) Variational total energies from $\Phi$- and $\Psi$-derivable theories. Int J Mod Phys B 13(05n06):535–541. http://www.worldscientific.com/doi/abs/10.1142/S0217979299000436

Arrachea L, Mucciolo ER, Chamon C, Capaz RB (2012) Microscopic model of a phononic refrigerator. Phys Rev B 86(12):125424. https://link.aps.org/doi/10.1103/PhysRevB.86.125424, arXiv:1203.2561v2

Baym G (1962) Self-consistent approximations in many-body systems. Phys Rev 127(4):1391–1401. http://prola.aps.org/abstract/PR/v127/i4/p1391_1; https://link.aps.org/doi/10.1103/PhysRev.127.1391

Baym G, Kadanoff LP (1961) Conservation laws and correlation functions. Phys Rev 124(2):287–299. http://prola.aps.org/abstract/PR/v124/i2/p287_1; https://link.aps.org/doi/10.1103/PhysRev.124.287

Bloch I, Zwerger W (2008) Many-body physics with ultracold gases. Rev Mod Phys 80(3):885–964. http://link.aps.org/doi/10.1103/RevModPhys.80.885

Bonitz M (2016) Quantum kinetic theory. Springer International Publishing, Cham. http://link.springer.com/10.1007/978-3-319-24121-0

Boström EV, Mikkelsen A, Verdozzi C, Perfetto E, Stefanucci G (2018) Charge separation in donor–C 60 complexes with real-time green functions: the importance of nonlocal correlations. Nano Lett 18(2):785–792. http://pubs.acs.org/doi/10.1021/acs.nanolett.7b03995

Bruus H, Flensberg K (2004) Introduction to many-body quantum theory in condensed matter physics. Oxford University Press, Oxford

Cuevas JC, Scheer E (2010) Molecular electronics. World Scientific Publishing, Singapore

Dahlen N, van Leeuwen R (2007) Solving the Kadanoff-Baym equations for inhomogeneous systems: application to atoms and molecules. Phys Rev Lett 98(15):153004. http://link.aps.org/doi/10.1103/PhysRevLett.98.153004

Datta S (2005) Quantum transport: atom to transistor. https://doi.org/10.1017/CBO9781139164313; http://books.google.com/books?hl=en&lr=&id=Yj50EJoS224C&oi=fnd&pg=PR9&dq=Quantum+Transport+:+Atom+to+Transistor&ots=jmVfovCmEu&sig=oYLWTfZuNxd44-WwK4FwD-Tmg98

De Dominicis C (1963) Variational statistical mechanics in terms of "Observables" for normal and superfluid systems. J Math Phys 4(2):255–265. http://aip.scitation.org/doi/10.1063/1.1703949

Fetter AL, Walecka JD (2003) Quantum theory of many-particle theory. Dover Publications, New York

Flick J, Ruggenthaler M, Appel H, Rubio A (2017) Atoms and molecules in cavities, from weak to strong coupling in quantum-electrodynamics (QED) chemistry. Proc Natl Acad Sci 114(12):3026–3034. http://www.pnas.org/lookup/doi/10.1073/pnas.1615509114

Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139:A796. http://link.aps.org/doi/10.1103/PhysRev.139.A796

Hedin L, Lundqvist S (1970) Effects of electron-electron and electron-phonon interactions on the one-electron states of solids. Solid State Phys 23:1–181. https://doi.org/10.1016/S0081-1947(08)60615-3; http://linkinghub.elsevier.com/retrieve/pii/S0081194708606153

Henneberger K, Moldzio U, Güldner H (2000) Correlation effects and quantum kinetics in pulse excited semiconductors. In: Progress in Nonequilibrium Green's Functions. World Scientific, pp 180–211. https://doi.org/10.1142/9789812793812_0015; http://www.worldscientific.com/doi/abs/10.1142/9789812793812_0015

Karlsson D, van Leeuwen R (2016) Partial self-consistency and analyticity in many-body perturbation theory: particle number conservation and a generalized sum rule. Phys Rev B 94(12):125124. https://link.aps.org/doi/10.1103/PhysRevB.94.125124

Khosravi E, Uimonen AM, Stan A, Stefanucci G, Kurth S, van Leeuwen R, Gross EKU (2012) Correlation effects in bistability at the nanoscale: steady state and beyond. Phys Rev B 85(7):075103. http://link.aps.org/doi/10.1103/PhysRevB.85.075103

Latini S, Perfetto E, Uimonen AM, van Leeuwen R, Stefanucci G (2014) Charge dynamics in molecular junctions: nonequilibrium Green's function approach made fast. Phys Rev B 89(7):075306. http://link.aps.org/doi/10.1103/PhysRevB.89.075306

Lipavský P, Špička V, Velický B (1986) Generalized Kadanoff-Baym ansatz for deriving quantum transport equations. Phys Rev B 34(10):6933–6942. http://prb.aps.org/abstract/PRB/v34/i10/p6933_1

Mahfouzi F, Nikolić BK (2014) Signatures of electron-magnon interaction in charge and spin currents through magnetic tunnel junctions: a nonequilibrium many-body perturbation theory approach. Phys Rev B 90(4):045115. https://link.aps.org/doi/10.1103/PhysRevB.90.045115

Martin PC, Schwinger J (1959) Theory of many-particle systems. I. Phys Rev 115(6):1342–1373. https://link.aps.org/doi/10.1103/PhysRev.115.1342

de Melo PMMC, Marini A (2016) Unified theory of quantized electrons, phonons, and photons out of equilibrium: A simplified ab initio approach based on the generalized Baym-Kadanoff ansatz. Phys Rev B 93(15):155102. https://link.aps.org/doi/10.1103/PhysRevB.93.155102

Myöhänen P, Stan A, Stefanucci G, van Leeuwen R (2008) A many-body approach to quantum transport dynamics: initial correlations and memory effects. Europhysics Lett

84(6):67001. https://doi.org/10.1209/0295-5075/84/67001; http://stacks.iop.org/0295-5075/84/i=6/a=67001?key=crossref.7f76c45bb53d32a8d0f63a9c53ac212b

Myöhänen P, Stan A, Stefanucci G, van Leeuwen R (2009) Kadanoff-Baym approach to quantum transport through interacting nanoscale systems: From the transient to the steady-state regime. Phys Rev B 80(11):115,107. https://doi.org/10.1103/PhysRevB.80.115107, URL http://prb.aps.org/abstract/PRB/v80/i11/e115107 https://link.aps.org/doi/10.1103/PhysRevB.80.115107

Myöhänen P, Tuovinen R, Korhonen T, Stefanucci G, van Leeuwen R (2012) Image charge dynamics in time-dependent quantum transport. Phys Rev B 85(7):075105. http://prb.aps.org/abstract/PRB/v85/i7/e075105; https://link.aps.org/doi/10.1103/PhysRevB.85.075105

Perfetto E, Uimonen AM, van Leeuwen R, Stefanucci G (2015) First-principles nonequilibrium Green's-function approach to transient photoabsorption: application to atoms. Phys Rev A 92(3):033419. http://link.aps.org/doi/10.1103/PhysRevA.92.033419

Perfetto E, Sangalli D, Marini A, Stefanucci G (2018) Ultrafast charge migration in XUV photoexcited phenylalanine: a first-principles study based on real-time nonequilibrium Green's functions. J Phys Chem Lett 9(6):1353–1358. http://pubs.acs.org/doi/10.1021/acs.jpclett.8b00025

Puig von Friesen M, Verdozzi C, Almbladh CO (2010) Kadanoff-Baym dynamics of Hubbard clusters: performance of many-body schemes, correlation-induced damping and multiple steady and quasi-steady states. Phys Rev B 82(15):155108. http://link.aps.org/doi/10.1103/PhysRevB.82.155108

Ruggenthaler M, Flick J, Pellegrini C, Appel H, Tokatly IV, Rubio A (2014) Quantum-electrodynamical density-functional theory: bridging quantum optics and electronic-structure theory. Phys Rev A At Mol Opt Phys 90(1):1–26. https://doi.org/10.1103/PhysRevA.90.012508, 1403.5541

Säkkinen N (2016) Application of time-dependent many-body perturbation theory to excitation spectra of selected finite model systems. PhD thesis, University of Jyväskylä. http://urn.fi/URN:ISBN:978-951-39-6814-4

Säkkinen N, Peng Y, Appel H, van Leeuwen R (2015a) Many-body Green's function theory for electron-phonon interactions: ground state properties of the Holstein dimer. J Chem Phys 143(23):234101. https://doi.org/10.1063/1.4936142; http://aip.scitation.org/doi/10.1063/1.4936142, 1403.2968

Säkkinen N, Peng Y, Appel H, van Leeuwen R (2015b) Many-body Green's function theory for electron-phonon interactions: the Kadanoff-Baym approach to spectral properties of the Holstein dimer. J Chem Phys 143(23):234102. https://doi.org/10.1063/1.4936143 http://aip.scitation.org/doi/10.1063/1.4936143, 1507.04726

Schlünzen N, Hermanns S, Bonitz M, Verdozzi C (2016) Dynamics of strongly correlated fermions: Ab initio results for two and three dimensions. Phys Rev B 93(3):035107. https://link.aps.org/doi/10.1103/PhysRevB.93.035107

Schüler M, Berakdar J, Pavlyukh Y (2016) Time-dependent many-body treatment of electron-boson dynamics: application to plasmon-accompanied photoemission. Phys Rev B 93(5):054303. https://link.aps.org/doi/10.1103/PhysRevB.93.054303, 1510.08650

Sentef MA, Kemper AF, Georges A, Kollath C (2016) Theory of light-enhanced phonon-mediated superconductivity. Phys Rev B 93(14):144506. https://link.aps.org/doi/10.1103/PhysRevB.93.144506, 1505.07575

Špička V, Velický B, Kalvová A (2014) Electron systems out of equilibrium: nonequilibrium Green's function approach. Int J Mod Phys B 28(23):1430013. https://doi.org/10.1142/S0217979214300138; http://www.worldscientific.com/doi/abs/10.1142/S0217979214300138?journalCode=ijmpb&quickLinkVolume=28&quickLinkIssue=23&quickLinkPage=1430013&selectedTab=citation&volume=28#.VvAQ0PsV1sE.mendeley

Stan A, Dahlen NE, van Leeuwen R (2009) Time propagation of the Kadanoff-Baym equations for inhomogeneous systems. J Chem Phys 130(22):224101. https://doi.org/10.1063/1.3127247; http://www.ncbi.nlm.nih.gov/pubmed/19530756

Stefanucci G, Almbladh CO (2004) Time-dependent partition-free approach in resonant tunneling systems. Phys Rev B 69(19):195318. https://link.aps.org/doi/10.1103/PhysRevB.69.195318

Stefanucci G, van Leeuwen R (2013) Nonequilibrium many-body theory of quantum systems: a modern introduction. Cambridge University Press, Cambridge

Tuovinen R, Säkkinen N, Karlsson D, Stefanucci G, van Leeuwen R (2016) Phononic heat transport in the transient regime: an analytic solution. Phys Rev B 93(21):214301. http://arxiv.org/abs/1604.02298; http://link.aps.org/doi/10.1103/PhysRevB.93.214301

Uimonen AM, Khosravi E, Stefanucci G, Kurth S, van Leeuwen R, Gross EKU (2010) Real-time switching between multiple steady-states in quantum transport. J Phys Conf Ser 220(1):012018. https://doi.org/10.1088/1742-6596/220/1/012018; http://stacks.iop.org/1742-6596/220/i=1/a=012018

Uimonen AM, Khosravi E, Stan A, Stefanucci G, Kurth S, van Leeuwen R, Gross EKU (2011) Comparative study of many-body perturbation theory and time-dependent density functional theory in the out-of-equilibrium Anderson model. Phys Rev B 84(11):115103. http://prb.aps.org/abstract/PRB/v84/i11/e115103; https://link.aps.org/doi/10.1103/PhysRevB.84.115103

Wang JS, Agarwalla BK, Li H, Thingna J (2014) Nonequilibrium Green's function method for quantum thermal transport. Front Phys 9(6):673–697. http://link.springer.com/10.1007/s11467-013-0340-x, 1303.7317

Wick GC (1950) The evaluation of the collision matrix. Phys Rev 80(2):268–272. https://link.aps.org/doi/10.1103/PhysRev.80.268

# Non-equilibrium Dynamical Mean-Field Theory

## 18

Martin Eckstein

## Contents

### Abstract

Intense and ultrashort light pulses allow to investigate new states of matter in complex materials under non-equilibrium conditions. The formulation of dynamical mean-field theory (DMFT) using Keldysh Green's functions provides a framework to calculate the electronic structure of correlated materials out of equilibrium. The approach has contributed insight into a wide range of topics, including photo-induced processes in Mott insulators, non-equilibrium steady

M. Eckstein (✉)
Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

Theory Department: Theory of Correlated Systems out of Equilibrium, Max-Planck-Institut f. Struktur u. Dynamik der Materie, Hamburg, Germany
e-mail: martin.eckstein@fau.de

states in driven materials, and the fundamental question how isolated quantum systems thermalize. In this chapter we outline the theoretical foundations of non-equilibrium DMFT, present some of the major results so far, and briefly discuss future directions, which are needed in order to develop a framework in which material properties out of equilibrium can be obtained from first principles.

# 1    Introduction

Traditionally, computational material research is concerned with the equilibrium properties of matter. This is challenged by the new possibilities offered by ultrafast laser techniques (Giannetti et al. 2016). Intense femtosecond light pulses can stimulate new quantum states of matter which exist only out of equilibrium. Examples include possible light-induced superconductivity (Fausti et al. 2011) and the transient control of lattice constants and electronic properties through the excitation of an-harmonically coupled phonons (Rini et al. 2007). Photo-induced changes of material properties can even be long-lived, as evidenced by the switching to hidden phases (Ichikawa et al. 2011; Stojchevska et al. 2014) with new types of spin and orbital order. This new research direction poses great challenges to theory, as many fundamental concepts of condensed matter physics rely on the notion of thermal equilibrium or a projection to low-energy states. Even the single-particle band structure is in part an equilibrium concept: in complex materials, electronic quasiparticles are only well-defined close to the ground state, and the screened potential is determined by the collective response of the electronic system itself. To identify the theoretical challenges, let us discuss some questions in more detail.

**Transient nonthermal dynamics** For many situations one can assume rapid thermalization of electrons, so that photo-excited states in solids can be described in terms of an electronic quasi-equilibrium state. However, nontrivial transient phenomena such as the enhancement of electronic orders cannot be explained by a photo-induced hot electron temperature and must occur prior to thermalization. There is no generic answer to the question how fast a system of interacting particles would thermalize after a perturbation, although this concerns the basic foundations of statistical physics. Isolated model systems can behave strictly non-ergodic, in particular at integrability (Polkovnikov et al. 2011). Condensed matter systems are never ideal in that sense, but their short-time behavior may reflect the non-ergodic behavior of a related integrable system, leading to a two-stage relaxation with an earlier pre-thermal state (Moeckel and Kehrein 2008; Kollar et al. 2011), which can show long-range order even if the corresponding thermal state does not (Sciolla and Biroli 2013; Tsuji et al. 2013). Even far from integrable points, fast electronic thermalization is not obvious, e.g., close to the Mott transition (Sayyad and Eckstein 2016). The validity of Boltzmann kinetic equations, which can describe the dynamics in semiconductors (Haug and Jauho 2008), relies on the existence of well-defined quasiparticles and is not

clear for the ultrafast dynamics in strongly correlated systems. Nonthermal states are also important when non-perturbative electric fields act on a material. Short laser pulses can have peak fields that would break a material in the steady state, but not as long as the dynamics remains quantum coherent (Higuchi et al. 2017), when instead phenomena like Bloch oscillations and high-harmonic generation can occur.

**Non-equilibrium steady states** When a system is subject to a rapidly oscillating time-periodic perturbation, averaging over the fast driving can yield a new effective Hamiltonian, in analogy to dynamical stabilization of new equilibria in mechanical systems (Bukov et al. 2015). This so-called Floquet engineering is used to design tight-binding Hamiltonians in cold-gas experiments (Goldman and Dalibard 2014), including topologically nontrivial band structures (Jotzu et al. 2014). Floquet engineering can be used to manipulate emergent many-body interactions such as spin exchange interactions (Mentink et al. 2015) or superconducting pairing (Knap et al. 2016), which suggests many routes for material design out of equilibrium. The fundamental question in this context is whether a non-equilibrium system can be directed into the ground state or a thermal state of such engineered Hamiltonians or whether the process is dominated by heating effects. To answer this question, one needs to understand the non-equilibrium steady states in strongly correlated open quantum system, in which energy input and dissipation to the environment are balanced.

A viable approach to describe both the ultrafast transient evolution and non-equilibrium steady states of many-body systems is the Keldysh formalism. It is based on non-equilibrium Green's functions, which describe the propagation of particles and holes between different space-time points. Green's functions contain both the information on the spectrum (electronic structure, quasiparticle energies) and the occupation of states, which mutually depend on each other out of equilibrium. In equilibrium, dynamical mean-field theory (DMFT) nowadays provides a standard approach to describe correlated systems (Georges et al. 1996). DMFT maps a lattice model with local interactions to an effective impurity model, which can be solved numerically. This approximation accurately treats local temporal fluctuations and can therefore capture many phenomena that originate from the competition of atomic correlations and the itinerant behavior of electrons. DMFT can be applied to generic models obtained from density-functional theory and thus quantitatively predict the electronic structure of complex materials (Pavarini et al. 2014). When reformulated within the Keldysh formalism, DMFT can describe non-equilibrium phenomena in correlated materials (Aoki et al. 2014; Schmidt and Monien 2002; Freericks et al. 2006). The solution of the DMFT equations in real time still poses many numerical challenges. The purpose of this chapter is to explain the foundations of non-equilibrium DMFT within the Keldysh framework (Sect. 2), to discuss DMFT and the impurity problem out of equilibrium (Sect. 3), and to give examples for the application of non-equilibrium DMFT (Sect. 4). Current developments of non-equilibrium DMFT are briefly sketched in the outlook (Sect. 5).

## 2        Theoretical Foundations

This section summarizes theoretical aspects needed for the development of non-equilibrium DMFT, in particular the Keldysh formalism. Only the main concepts are presented. For more details we refer to the literature, e.g., books by Haug and Jauho (2008) and Kamenev (2011).

### 2.1        Models

We first define typical models which are addressed within non-equilibrium DMFT. The basic model for correlated electrons is the one-band Hubbard model,

$$H = \sum_{i,j,\sigma} v_{ij}\, c_{i\sigma}^{\dagger} c_{j\sigma} + U \sum_j n_{j\uparrow} n_{j\downarrow}. \tag{1}$$

Here $n_{j\sigma} = c_{j\sigma}^{\dagger} c_{j\sigma}$, $c_{j\sigma}^{\dagger}$ ($c_{j\sigma}$) create (annihilate) an electron with spin $\sigma$ at site $j$ of a crystal lattice, $v_{ij}$ is the hopping matrix element, and electrons interact via a local Coulomb interaction $U$. DMFT can also be formulated for models with more than one orbital per site and different *local* interactions, such as a coupling to localized spins (Kondo lattice model) or Einstein phonons (Holstein model). For equilibrium simulations, such lattice Hamiltonians can be downfolded from an ab initio band structure. Regarding non-equilibrium DMFT, the development of analogous downfolding schemes is still subject to current research.

To incorporate electric fields, a tight-binding model such as (1) is modified to

$$H_0 = \sum_{i,j,\sigma} v_{ij} e^{ieA_{ij}} c_{i\sigma}^{\dagger} c_{j\sigma} + \sum_{i\sigma} e\phi_i c_{i\sigma}^{\dagger} c_{i\sigma}, \tag{2}$$

where $e$ is the charge, $\phi_i = \phi(\boldsymbol{R}_i)$ is the scalar potential, and $A_{ij} = \int_{\boldsymbol{R}_i}^{\boldsymbol{R}_j} d\boldsymbol{r} \cdot \boldsymbol{A}(\boldsymbol{r})$ is a vector potential along a link of the lattice. This so-called Peierls substitution, which can be derived assuming localized Wannier orbitals, is the minimal extension of the lattice model that ensures a local gauge symmetry. It does not include terms such as Zeeman fields, Stark-shifts of orbitals, etc. Such terms may be described by additional free parameters in the noninteracting part of the Hamiltonian and do not change the general DMFT framework.

### 2.2        Green's Functions and Electronic Structure

In equilibrium, the electronic structure is fully characterized by the spectral function, which defines the electronic bands in an interacting system. A suitable generalization to non-equilibrium is given in terms of two-time Green's functions. One can start from the electron and hole propagators,

$$G^<_{a,a'}(t, t') = +i \langle c^\dagger_{a'}(t') c_a(t) \rangle, \quad G^>_{a,a'}(t, t') = -i \langle c_{a'}(t) c^\dagger_a(t') \rangle. \tag{3}$$

Here $c^\dagger_a$ and $c_a$ denote the creation and annihilation operators for an electron in a single-particle orbital $|a\rangle$, i.e., momentum or position and spin and orbital degrees of freedom. The time-dependence of the operators is understood in the Heisenberg picture, and $\langle \cdots \rangle = \mathrm{tr}[\rho_0 \cdots]$ averages over the density matrix of the initial equilibrium state at some early time $t_0$. To keep the notation concise, we will only indicate a diagonal momentum index $\boldsymbol{k}$ in the following, $G_{\boldsymbol{k}} \equiv G_{\boldsymbol{k},\boldsymbol{k}}$. The functions $G^<_{\boldsymbol{k}}(t, t')$ and $G^>_{\boldsymbol{k}}(t, t')$ give the amplitude for the propagation of a hole or an additional electron in the (possibly non-equilibrium) many-body state and therefore fully characterize the single-electron dynamics in the solid. In equilibrium, $G^<_{\boldsymbol{k}}(t, t')$ and $G^>_{\boldsymbol{k}}(t, t')$ depend only on time-difference $t - t'$. Using an expansion in many-body eigenstates (Lehmann representation), one can show that the Fourier transform $G_{\boldsymbol{k}}(\omega) = \int dt \, e^{i\omega t} G_{\boldsymbol{k}}(t, 0)$ satisfies the relations

$$G^<_{\boldsymbol{k}}(\omega) = 2\pi i \, A_{\boldsymbol{k}}(\omega) f(\omega), \quad G^>_{\boldsymbol{k}}(\omega) = -2\pi i \, A_{\boldsymbol{k}}(\omega)[1 - f(\omega)], \tag{4}$$

where $f(\omega)$ is the Fermi function and $A_{\boldsymbol{k}}(\omega)$ is the spectral function, which contains the information about the quasiparticle energies and their lifetimes.

### 2.2.1   Wigner Representation and Photoemission Spectrum

Equation (4) is an example of the fluctuation dissipation theorem and allows to rigorously separate spectrum and occupations in equilibrium. Out of equilibrium, such a separation is not possible in general. When time-translational invariance is lost, a suitable representation for two-time functions $F(t, t')$ is the Wigner transform, where one introduces average time $t_{\mathrm{av}} = (t + t')/2$ and relative time $t_{\mathrm{rel}} = t - t'$ and Fourier-transforms with respect to $t_{\mathrm{rel}}$,

$$F(t_{\mathrm{av}}, \omega) = \int dt_{\mathrm{rel}} \, e^{i\omega t_{\mathrm{rel}}} \, F(t_{\mathrm{av}} + t_{\mathrm{rel}}/2, t_{\mathrm{av}} - t_{\mathrm{rel}}/2). \tag{5}$$

The Wigner transform $G^<(t, \omega)$ has an intuitive interpretation: the factorization of $-i G^<_{\boldsymbol{k}}(\omega)$ as product of a density of states and an occupation function (cf. Eq. (4)) suggests that the Wigner transform $G^<_{\boldsymbol{k}}(t, \omega)$ gives the probability distribution to remove a particle with energy $\hbar\omega$ and momentum $\boldsymbol{k}$ from the system at time $t$, e.g., in a time- and angle-resolved photoemission experiments. This interpretation is true, when $G^<_{\boldsymbol{k}}(t, \omega)$ is averaged over a time and frequency window which satisfies the uncertainty relation $\Delta t \Delta \omega \gtrsim 1$, in analogy to the Wigner phase-space density in semiclassical physics, which becomes of positive phase-space density positive after a suitable average of position and momentum over a phase-space volume $\hbar$.

More precisely, time-resolved photoemission measures the probability that an electron with energy $E_{\mathrm{out}}$ is emitted under the action of a probe pulse with finite duration and given delay $t_p$ with respect to some excitation ("time zero"). The

signal can be obtained using time-dependent perturbation theory in the light-matter coupling (Freericks et al. 2009; Eckstein and Kollar 2008),

$$I(E, t_p) \propto \int dt\, dt'\; e^{iE(t-t')}(-i)G^<(t, t')\; S(t - t_p)S(t' - t_p)^*. \tag{6}$$

Here the probe pulse has a frequency $\Omega$ and an envelope $S(t)$, and $E = (E_{\text{out}} - \Omega)$ is the energy extracted from the solid. Matrix elements, which would select out certain orbital and momentum components of $G$, are not shown in Eq. (6). In addition, the sudden approximation was employed in the derivation, which assumes that there is no interaction between the outgoing electrons and the solid. Using a Gaussian profile $S(t) = \exp\!\left(-t^2/2\Delta t^2\right)$ with duration $\Delta t$, Eq. (6) transforms to the convolution of $G^<(\omega, t)$ with the kernel $e^{-\frac{(t-t-p)^2}{\Delta t^2}} e^{-(\omega-E)^2\Delta t^2}$, which implies the abovementioned probabilistic interpretation of non-equilibrium Green's functions.

## 2.3 Keldysh Formalism: Real-Time Evolution

In order to describe the time evolution of a quantum system, we aim to compute observables of the general form

$$\langle \mathcal{O}(t_1)\rangle = \text{tr}\!\left[\rho_0\, \mathcal{U}(t_1, t_0)^\dagger\, \mathcal{O}\, \mathcal{U}(t_1, t_0)\right]. \tag{7}$$

Here $\rho_0$ is the density matrix which defines the state at some initial time $t_0$, and $\mathcal{U}(t, t_0)$ is the unitary time evolution operator, which is given by the time-ordered exponential $T_t \exp\!\left[-i\int_{t_0}^t d\bar{t}\, H(\bar{t})\right]$. Equation (7) describes the evolution of an isolated quantum system without contact to environment.

As a mathematical trick, the time-ordering in $\mathcal{U}(t, t_0)$ and the corresponding anti-time-ordering in $\mathcal{U}(t, t_0)^\dagger$ can be combined into an ordering along a time contour which extends from $t_0$ forward and then backward in time. Furthermore, the thermal density matrix can be written as a time evolution operator along an imaginary time axis $[0, -i\beta]$. The three branches can be combined into a single $L$-shaped contour $\mathscr{C}$, as depicted in Fig. 1. Together with $\mathscr{C}$ we introduce the contour-ordering operator
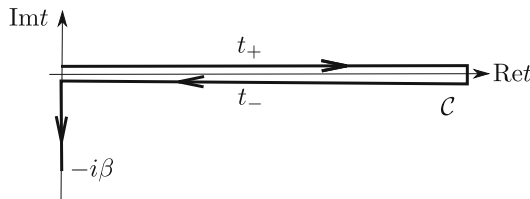


**Fig. 1** The $L$-shaped Keldysh contour $\mathscr{C}$. Times on the horizontal branches are purely real and denoted with a book-keeping index $\pm$ to distinguish upper and lower branch, respectively. The arrows denote the time-ordering along $\mathscr{C}$ from earlier to later contour times

$$T_{\mathscr{C}}A(t)B(t') = \begin{cases} A(t)B(t') & \text{if } t \text{ later on } \mathscr{C} \text{ than } t' \\ \xi B(t')A(t) & \text{if } t' \text{ later on } \mathscr{C} \text{ than } t \end{cases}, \tag{8}$$

where the sign $\xi$ is $-1$ $(+1)$ if the permutation of $A$ and $B$ involves an odd (even) number of permutations of fermion creation or annihilation operators. The expectation value (7) can then be written as

$$\langle O(t) \rangle = \frac{1}{Z} \text{tr}[T_{\mathscr{C}} e^{-i \int_{\mathscr{C}} d\bar{t} H(\bar{t})} O(t_+)], \tag{9}$$

where $\int_{\mathscr{C}} dt$ is the integral along the contour. The contour ordering allows to formally extend the Matsubara formalism for many-body systems to real time. Diagrammatic perturbation theory and path integrals can be reformulated by merely replacing integrals over imaginary time by integrals over $\mathscr{C}$: in particular, the path-integral action for a generic Hamiltonian $H$ is given by

$$S = \int_{\mathscr{C}} dt \sum_a c_a^*(t) i \partial_t c_a(t) - H[c^*(t), c(t)], \tag{10}$$

the partition function can be written as a path integral $Z = \int \mathscr{D}[c^*c] e^{iS}$, and the expectation value (9) becomes an average $\langle O(t) \rangle_S$, with $\langle \cdots \rangle_S = Z^{-1} \int \mathscr{D}[c^*c] e^{iS} \cdots$. The electron and hole propagators (3) appear naturally as components of a single contour-ordered Green's function,

$$G_{ab}(t, t') = -i \langle c_a(t) c_b^*(t') \rangle_S, \tag{11}$$

with $G(t_+, t'_-) = G^<(t, t')$ and $G(t_-, t'_+) = G^>(t, t')$ for $t, t' \in \mathbb{R}$.

The analogy to the Matsubara formalism implies that contour-ordered Green's functions satisfy Wick's theorem when the action is quadratic (the latter is merely a consequence of Gaussian integration). Hence the rules for constructing Feynman diagrams are identical to the Matsubara formalism (Mahan 2000). We can introduce the two-time self-energy $\Sigma(t, t')$ as the sum over all irreducible diagrams. The Dyson equation then relates interacting ($G$) and noninteracting ($G_0$) Green's functions,

$$G = G_0 + G_0 * \Sigma * G = G_0 + G * \Sigma * G_0. \tag{12}$$

Here $*$ denotes convolution over $\mathscr{C}$. In contrast to equilibrium, where a solution of the Dyson equation is obtained in Fourier space, the Dyson equation in real time is an integral equation on $\mathscr{C}$. Its solution presents a formidable numerical task, which is discussed in the next section.

## 2.4 Kadanoff-Baym Equations

Given an approximation to the self-energy, the determination of the Green's function requires the solution of the Dyson equation (12). With a suitable time grid of $M$ time slices along $\mathscr{C}$, the linear integral equation (12) may be reduced to a matrix equation of dimension $M$ (Freericks 2008). In practice, however, one mostly proceeds differently, in order to make use of the causality in the time-propagation: on a given time grid, the solution of Eq. (12) can be extended from real times $t, t' \leq n\Delta t$ to the domain $t, t' \leq (n + 1)\Delta t$ without modification of the solution on previous times. The Dyson equation can therefore be formulated as a time-propagation scheme. This is useful because typically $\Sigma[G]$ itself is a (causal) functional of the Green's function, e.g., in the form of a perturbation expression. Mathematically, the causal structure implies that integral equations on $\mathscr{C}$ can be reduced to a set of Volterra equations, which can be solved using high-order accurate algorithms (Brunner and van der Houwen 1986). The representation of the Dyson equation in terms of coupled integral equations for the real-time propagators is referred to as Kadanoff-Baym equations. Various implementations have been presented, which differ in the representation of the Green's function in terms of the lesser, Keldysh, and retarded components (Bonitz 2000; Eckstein et al. 2010a; Aoki et al. 2014).

Applications of two-time Green's functions to condensed matter systems range from the dynamics of screening in semiconductors (Bányai et al. 1998) to electron-phonon coupled superconductors (Kemper et al. 2017). In general, for $M$ time slices, the required computational resources scale like $\mathscr{O}(M^2)$ for memory and $\mathscr{O}(M^3)$ for CPU time. This is a major limiting factor, in particular when Green's functions carry many orbital indices. Simulations based on full Kadanoff-Baym equations have therefore been restricted to few-orbital model systems. An important step toward first-principle simulations of non-equilibrium electronic structure would be a formulation of the GW approximation (Onida et al. 2002) in the time domain. Within GW, the self-energy $\Sigma_{\boldsymbol{k}}$ is expanded to leading order in the screened interaction $W_{\boldsymbol{k}}$, and the polarizability $\chi_{\boldsymbol{k}}$ is in turn computed using the Lindhard function. This captures nonlocal phenomena resulting from charge fluctuations, including screening and plasmonic modes. Even though the Dyson equations for $W_{\boldsymbol{k}}$ and $\Sigma_{\boldsymbol{k}}$ can be computed in parallel, the calculation of the self-energies and polarization functions requires a summation over all $\boldsymbol{k}$-points, which makes the solution of the full lattice GW simulations a formidable task. Results for a four-band system have been discussed in Golež et al. (2016).

The numerical cost of solving the Kadanoff-Baym equations motivates a possible truncation of the integrals. The convolution $\Sigma * G$ in the Dyson equation corresponds to memory effects in the propagation of the single-particle Green's functions. In contrast, kinetic equations provide a differential equation for the propagation of a single-particle density matrix (occupations) without memory integrals. A possible approximation to the Kadanoff-Baym approximation which focuses on the single-particle density matrix is the so-called generalized Kadanoff-Baym ansatz (GKBA) (Lipavský et al. 1986), used in combination with a mean-field approximation to the retarded propagators, i.e., assuming ideal quasiparticles. To what extent the GKBA

is useful in condensed matter systems relies on the validity of the quasiparticle approximation, which needs to be critically accessed for correlated systems.

## 3    The Dynamical Mean-Field Formalism

### 3.1    Self-Consistent Mapping to an Impurity Model

The starting point for the development of DMFT has been the limit of infinite lattice coordination number (Metzner and Vollhardt 1989). For a fermionic tight-binding model, this limit is defined together with a rescaling of the hopping matrix elements $v$, in order to preserve the competition between kinetic and interaction energies which underlies many phenomena in correlated electron systems. For a $d$-dimensional cubic lattice with nearest neighbor hopping, the correct rescaling is $v = v_*/\sqrt{2d}$, where $v_*$ is kept constant for $d \to \infty$. The infinite-dimensional limit implies that the self-energy becomes momentum-independent, i.e., local in real-space (Müller-Hartmann 1989),

$$\Sigma_{ij}(t, t') = \delta_{ij} \, \Sigma_{ii}(t, t'). \tag{13}$$

This fact is proven by counting powers of $d$ in the diagrams and therefore holds equally well in the Keldysh and Matsubara formalism.

The locality of the self-energy can be taken as a non-perturbative approximation for finite-dimensional systems. The summation of local $\Sigma$-diagrams to infinite order is facilitated by the solution of an auxiliary impurity model (Georges and Kotliar 1992), as illustrated in Fig. 2. The mapping to the impurity model can be explained, e.g., in a functional language: in general, the self-energy can be expressed as a functional of the fully interacting Green's function. This so-called skeleton functional $\Sigma^{\text{skel}}[G]$ consists of the sum over all self-energy diagrams in which the lines do not have self-energy insertions and are in turn replaced by the fully
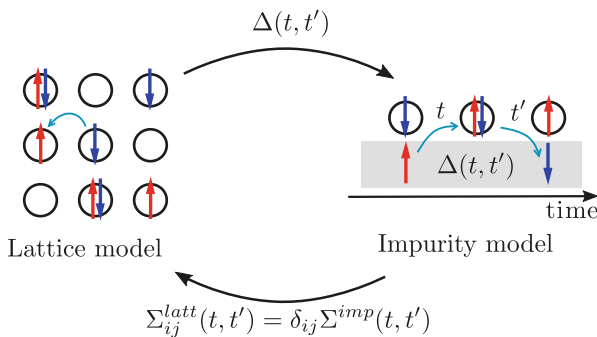


**Fig. 2** Mapping from a lattice model with local self-energy to an impurity model with hybridization function $\Delta(t, t')$

interacting $G$. For an auxiliary impurity problem in which one site of the lattice is embedded in a noninteracting medium, the skeleton functional $\Sigma^{\text{imp}}[G]$ does not depend on the form of the noninteracting part of the Hamiltonian. The solution of the impurity model for the Green's function $G_{\text{imp}}$ and self-energy $\Sigma_{\text{imp}}$ can therefore be seen as a device to approximate the lattice functional $\Sigma^{\text{skel}}_{jj}[G]$ with the contributions from the local Green's function $G_{jj}$ to all orders. This argument was first given for the Falicov-Kimball model by Brandt and Mielsch (1991). An alternative intuitive and elegant way to formulate the mapping to an impurity model is given by the so-called cavity method, as introduced in Georges et al. (1996).

For the single-band Hubbard model, the corresponding impurity model has the form of a general Anderson impurity Hamiltonian (Georges and Kotliar 1992)

$$H_{\text{imp}} = U c^\dagger_\uparrow c_\uparrow c^\dagger_\downarrow c_\downarrow + \epsilon_f \sum_\sigma c^\dagger_\sigma c_\sigma + \sum_{p\sigma}(V_p(t)c^\dagger_\sigma a_{p\sigma} + h.c.) + \sum_{p\sigma} \epsilon_p(t) a^\dagger_{p\sigma} a_{p\sigma},$$
(14)

where one interacting site is coupled to noninteracting bath orbitals $p$. Equivalently, we can write the model as a single site with a general quadratic action, which is defined through the hybridization function $\Delta(t, t')$,

$$S_{\text{imp}} = \int_{\mathscr{C}} dt \Big\{ \sum_\sigma c^*_\sigma(t) i \partial_t c_\sigma(t) - H_{\text{loc}}(t) \Big\} - \sum_\sigma \int_{\mathscr{C}} dt dt' \, c^*_\sigma(t) \Delta(t, t') c_\sigma(t').$$
(15)

The impurity Dyson equation $G_{\text{imp}} = \mathscr{G} + \mathscr{G} * \Sigma_{\text{imp}} * G_{\text{imp}}$ fixes a relation between the impurity Green's function $G_{\text{imp}}(t, t') = -i\langle c(t)c^*(t')\rangle_{S_{\text{imp}}}$ and the impurity self-energy ($\mathscr{G}[\Delta]$ is the noninteracting impurity Green's function). The impurity self-energy is then taken as an approximation for the local lattice self-energy,

$$\Sigma_{\text{imp}}(t, t') = \Sigma_{jj}(t, t'),$$
(16)

and the lattice Green's function $G_{ij}(t, t')$ is obtained from the lattice Dyson equation. (To simplify the notation, the framework is presented for a translationally invariant state, i.e., a possible dependence of local quantities on the site is not shown.) At this stage, the noninteracting lattice Green's function incorporates the effect of external electromagnetic fields. The DMFT equations are closed when the bath $\Delta(t, t')$ (or equivalently the bath parameters $V_p(t)$ and $\epsilon_p(t)$) are determined self-consistently such that the local lattice Green's function and the impurity Green's functions $G_{\text{imp}}(t, t') = G_{jj}(t, t')$ match.

In equilibrium, the self-consistent solution of the DMFT equations is achieved iteratively. One can start from a guess for $\Sigma$, solve the lattice Dyson equation to obtain $G_{jj}$, invert the impurity Dyson equation to get $\Delta$, and solve the impurity model with action (15) to get $\Sigma$. DMFT is in essence a solution of the lattice Dyson equation with a nonlinear functional $\Sigma_{jj}[G]$ defined through the auxiliary

impurity model. Because this functional is causal, the DMFT equations on the Keldysh contour can be solved step by step in time, as explained for the Dyson equation in Sect. 2.4. The main conceptual and computational bottleneck in the DMFT framework is the solution of the impurity model, which we summarize in the following section.

## 3.2    Impurity Solvers

At present there is no general-purpose approach to solve the impurity problem, but a number of techniques which work well in certain parameter regimes, or provide exact results in a short time.

### 3.2.1    Weak-Coupling Expansions

When perturbation theory is applied to the impurity model, i.e., in combination with a self-consistent determination of the bath, it can capture non-perturbative phenomena like the Mott metal-insulator phase transition (Georges et al. 1996). The most important example is iterated perturbation theory (IPT), i.e., a second-order expansion of $\Sigma$ in $\mathscr{G}[\Delta]$. IPT is numerically relatively inexpensive (using IPT, the numerical cost of DMFT is determined by the solution of the lattice Dyson equation). Interesting applications include the study of nonthermal critical points in the quench dynamics of the antiferromagnetic spin-density wave (Tsuji et al. 2013). However, IPT is a non-conserving approximation. Diagrammatic approximations to the self-energy respect conservation laws like energy and particle number when they are based on a truncation of the skeleton expansion in terms of the interacting $G$ (Baym and Kadanoff 1961). In equilibrium, expansions in the *bare* Green's functions are often quantitatively more accurate, but their non-conserving nature can become problematic in the real-time evolution (Eckstein et al. 2010a). The use of IPT has so far remained restricted to the weak-coupling regime, although in equilibrium IPT extrapolates between weak and strong coupling for the half-filled single-band model.

### 3.2.2    Strong-Coupling Expansions

The perturbative expansion in the hybridization function is a flexible approach to study systems in the strongly interacting regime. Starting from the isolated impurity Hamiltonian $H_{\text{loc}}$, it defines propagators $\mathscr{G}_{a,a'}$ in terms of the *many-body* Fock states of the impurity site. Particle creation and annihilation events on the impurity induce transitions between the states $a$, i.e., the hybridization $\Delta$ in the action (15) corresponds to a retarded "two-body" interaction between the many-body states. Although Wick's theorem is not valid when the zeroth order of the expansion corresponds to a non-quadratic Hamiltonian, one can re-sum the expansion of the partition function in terms of $\Delta$ to generate conserving approximations. The adaption of the strong-coupling techniques to the Keldysh contour is described in Eckstein and Werner (2010b). The general starting point of the expansion allows for an extension to arbitrary local Hamiltonians, including electron-phonon coupled

systems (Werner and Eckstein 2013), bosonic DMFT (Strand et al. 2015), or multi-orbital Hubbard models.

The approach has been widely used to study photo-induced dynamics in Mott insulators (Sect. 4.2). However, only the first order (the so-called non-crossing approximation, NCA) and to some extent the second order are numerically feasible. Higher-order expansions require multidimensional integrations over time, which become costly in particular for multi-orbital systems, when the matrix dimension of the propagators $\mathscr{G}_{a,a'}$ is set by the dimension of the local Fock-space. Low-order expansions cannot recover the weakly interacting limit, which has so far excluded studies of correlated metallic phases at low-energy.

### 3.2.3    Quantum Monte Carlo (QMC)

QMC corresponds to a stochastic summation of the perturbation expansion in the hybridization function or the interaction to all orders. This is nowadays the standard approach to solve DMFT in equilibrium at finite temperature (Gull et al. 2011). In real time, however, the terms of the perturbation expansion become complex-valued (dynamical sign problem), and Monte Carlo is so far limited to times of the order of few inverse hoppings (Mühlbacher and Rabani 2008; Werner et al. 2009). QMC solvers have been used to address fundamental problems related to the short-time dynamics (Sect. 4.1), but for the study of the photo-induced dynamics, few hopping times are usually not yet sufficient. A recent development to possibly overcome the dynamical sign problem has been presented by Cohen et al. (2015), using the causality of the time-propagation within a diagrammatic Monte Carlo approach. It will be interesting to apply this approach in the context of non-equilibrium DMFT.

### 3.2.4    Hamiltonian-Based Impurity Solvers

A different direction to develop numerically exact impurity solvers is based on exact diagonalization. Here one takes a finite system such as the single-impurity Anderson model Eq. (14), to approximately reproduce a given hybridization function $\Delta(t, t')$, and uses the size of the bath as a numerical control parameter. Finite representations of the bath have been derived both for the steady state, using an impurity model in which the bath-sites are coupled to dissipative Lindblad terms (Arrigoni et al. 2013) and for real time (Gramsch et al. 2013). In the latter case, wave-function propagation techniques based on matrix product states (MPS) can be used to compute the Green's function (Wolf et al. 2014). The numerical effort in these Hamiltonian-based techniques increases exponentially with the number of orbitals in the representation, which limits the frequency resolution in the steady state and the accessible times in the real-time formalism.

## 3.3    Non-equilibrium Steady States

The Green's function formalism can easily deal with open quantum systems in a quite general environment. When a system is subject to a steady perturbation and coupled to an environment, it can eventually reach a non-equilibrium steady state

in which energy input is balanced by dissipation. We can distinguish two different settings.

### 3.3.1 Time-Independent Perturbations

This includes in particular the coupling to external leads at a given voltage bias. For time-independent perturbations, we can expect that Green's functions depend only on the time-difference, but the Fermi function in (4) is replaced by a nonthermal distribution function. One can impose the time-translational invariance as an ansatz in the DMFT equations. The impurity model then has to be solved for a bath $\Delta$ representing a non-equilibrium steady state, which has so far been done using either IPT or the auxiliary Master equations (Arrigoni et al. 2013). The steady-state formalism has been used, e.g., to discuss the negative differential resistance in a correlated metal (Amaricci et al. 2012) or resistive switching at the Mott transition (Li et al. 2015).

### 3.3.2 Time-Periodic Perturbations

As discussed in the introduction, time-periodic perturbations offer the possibility to control the Hamiltonian parameters in a so-called Floquet Hamiltonian. Under the effect of steady driving, the system can approach a time-periodic steady state, in which all two-time correlation functions remain invariant if both arguments are shifted by the periodicity $\tau$ of the driving, $G^{>,<}(t + \tau, t' + \tau) = G^{>,<}(t, t')$. Again this can be used as an ansatz in the formalism, and Green's functions can be parametrized using a standard Fourier transform for the difference and a discrete Fourier transform for the average time. Within DMFT, this formalism has been pioneered by Schmidt and Monien (2002) and taken up by Tsuji et al. (2008) and Joura et al. (2008). For the details of the representation, we refer to Aoki et al. (2014). As one of the more recent results, e.g., the steady state of a Holstein model with parametric driving of the phonons was studied to investigate the possibility of light-enhanced superconductivity (Murakami et al. 2017).

## 4 Applications

In this section we present some applications of the non-equilibrium DMFT framework. We discuss results for the interaction quench in the Hubbard model, which have provided an important benchmark for the development of the theory, and results for the photo-doped Mott insulators, which best represent the path toward material modelling. This is not at all intended to be an exhaustive list of references. Among the many topics which are omitted here are in particular results for driven systems (Tsuji et al. 2011; Mentink et al. 2015; Murakami et al. 2017), inhomogeneous systems (Eckstein and Werner 2013a), the dielectric breakdown of Mott insulators (Eckstein et al. 2010b), or long-range interactions in extended DMFT (Golež et al. 2015).

## 4.1    Quenches and Thermalization

A fundamental question of statistical physics is whether and how an isolated many-body system thermalizes after an excitation. Motivated by experiments in artificial quantum systems, where parameters can be controlled to high-precision, this question has been investigated in a large class of models (Polkovnikov et al. 2011). Non-equilibrium DMFT opened the unique possibility for numerically exact studies in systems of dimension $d > 2$. Beyond the relevance for the understanding of thermalization, these studies serve as a useful benchmark results for the development of impurity solvers.

The DMFT equations were solved using an QMC impurity solver for a quench in the single-band Hubbard model, where the interaction is suddenly changed from zero to a finite value $U > 0$ (Eckstein et al. 2009). Corresponding numerically exact results for quenches from the atomic limit have been obtained with an MPS-based impurity solver (Balzer et al. 2015). The results for the quench from $U = 0$ are best understood by the evolution of the momentum occupation $n(\boldsymbol{k}) = \langle c_{\boldsymbol{k}}^{\dagger}(t) c_{\boldsymbol{k}}(t) \rangle$. In the initial state, taken at temperature $T = 0$, $n(\boldsymbol{k})$ shows a step discontinuity of size one at the Fermi energy ($\epsilon_{\boldsymbol{k}_F} = 0$). For quenches to small values of $U$, one observes a rapid relaxation to a nonthermal state in which an exact discontinuity at $\epsilon_{\boldsymbol{k}} = 0$ is retained. For quenches to large $U$, $n_{\boldsymbol{k}}$ shows damped collapse and revival oscillations, which reflect the behavior in the atomic limit, where the dynamics is perfectly $2\pi/U$-periodic. Within a sharp crossover region between large and small $U$, rapid thermalization is observed, i.e., within few hopping times the single-particle properties of the system can be described well by a system in thermal equilibrium with the same energy. The small $U$ behavior corresponds to pre-thermalization (Moeckel and Kehrein 2008), which is representative for the dynamics of near-integrable systems, and can be understood in terms of a perturbative time evolution with approximate constants of motion (Kollar et al. 2011). Apart from DMFT only exact numerics for finite or one-dimensional systems currently allow to address the breakdown of the perturbative evolution.

## 4.2    Photo-Doping in Mott Insulators

The Mott metal-insulator transition, at which electrons get localized by a strong Coulomb interaction, is one of the hallmarks of strong correlations in solids. It happens for a large class of transition metal compounds or molecular crystals with narrow conduction bands (Imada et al. 1998). In the Mott phase, spin and orbital degrees of freedom are still active, which is the origin of a variety of complex magnetic and orbital orders. The interaction of doped electrons or holes with such short-ranged magnetic and orbital correlations may furthermore lead to new phases such as high-temperature superconductivity. DMFT has been instrumental in the understanding of the Mott transition, and it is therefore natural to use non-equilibrium DMFT to study the wide range of phenomena related to photo-doped

Mott insulators. This section will outline some major results for the one-band Hubbard model.

**Thermalization of small-gap Mott insulators** DMFT predicts a first-order phase transition from a Fermi liquid to a Mott insulator around $U \approx W$ in the paramagnetic phase of the one-band Hubbard model, with a critical endpoint at a temperature $T^*$, above which the transition becomes a crossover. In the correlated metallic phase, the spectral function features a quasiparticle band around $\omega = 0$ in addition to Hubbard bands around $\omega = \pm U/2$, which corresponds to strongly renormalized quasiparticles. At the Mott transition, the quasiparticle band disappears, and a gap is opened. A straightforward excitation protocol is an electric field pulse with frequency $\Omega \approx U$, which generates charge excitations in the Mott insulator, i.e., doubly occupied and empty sites. This setting was analyzed in Eckstein and Werner (2011). Without coupling to environment (which is justified for times when lattice and electronic subsystem can be treated as decoupled), the system thermalizes on very short timescales if the interaction is close to the metal-insulator transition, and the final state is in the crossover regime. Thermalization can be verified, e.g., by considering the evolution of the double occupancy or the fluctuation dissipation relation (4). The rapid thermalization of small-gap Mott insulators is in agreement with very recent results for the relaxation of photo-doped $TaS_2$ (Ligges et al. 2018). The agreement of theory and experiment in this case is highly nontrivial, as the final thermalized states lie in the metal-insulator crossover regime where quasiparticles are not well-defined and standard kinetic approaches for the description electronic thermalization fail.

The thermalization of small-gap Mott insulators can be interpreted in terms of ultrafast impact ionization processes (Werner et al. 2014), where the kinetic energy of charge excitations is used to generate additional doublon-hole pairs, similar to an Auger process in atomic physics. Impact ionization is a carrier multiplication process which can potentially enhance the efficiency of light-harvesting devices. DMFT show that impact ionization processes in Mott insulators can be much faster than typical electron-phonon relaxation times, which would mostly dominate intra-band relaxation in semiconductors.

**Doublon-hole recombination** Deep in the Mott phase, the thermalization time increases exponentially with $U/W$. For $U \gg W$, this result has an intuitive interpretation: a single charge excitation can only take an energy comparable to the single-particle bandwidth $W$, while the conversion of a high-energy excitation into multiple excitations is generally expected to give exponentially long lifetimes. The decay of doublons has been discussed and measured with ultra-cold atoms (Strohmaier et al. 2010), and the DMFT results are in agreement with the observed exponential scaling (Eckstein and Werner 2011). In condensed matter systems, the decay of photo-excited carriers in Mott insulators can indeed range to picoseconds, i.e., thousands of hopping times (Okamoto et al. 2010). For a quantitative understanding, one must account for additional processes which involve spins (Lenarčič and Prelovšek 2014) or high energy phonons (Mitrano et al. 2014).

**The buildup of the Fermi liquid** The high temperature state after the thermalization of a correlated metal is a so-called bad-metal (Deng et al. 2013), in which coherent quasiparticles are absent and the scattering length becomes comparable to the lattice constant. A fundamental question is on which timescale the Fermi liquid can be reformed by cooling the electronic system, e.g., by coupling to phonons. Assuming rapid electronic thermalization, the corresponding timescale would be set by the rate of energy transfer to the bath. Surprisingly, there appears to be a much slower relaxation bottleneck of electronic origin at the onset of the Fermi liquid (Sayyad and Eckstein 2016). The finding implies that ultrashort quasiparticle lifetimes, i.e., rapid quasiparticle scattering, do not at all imply rapid thermalization, which clearly demonstrates the failure of naive kinetic descriptions. To date, the precise timescale for the formation of the Fermi liquid must be considered an open question. In practice, the absence of well-defined quasiparticles implies bad metallic properties of photo-doped Mott insulators (Eckstein and Werner 2013b), which could be analyzed using transient THz spectroscopy.

**Coupling of charge dynamics with antiferromagnetic correlations** At low temperatures, the half-filled single-band Hubbard model shows a tendency toward antiferromagnetism. In the Mott phase, antiferromagnetism is a consequence of a super-exchange interaction $4v_*^2/U$. Excitation of long-lived charge carriers, as described previously for the paramagnetic phase, now leads to both melting of the order parameter and a reduction of the kinetic energy of the carriers (Werner et al. 2012). This mutual interaction of the carriers with the spin background happens because a hole or doublon in the antiferromagnetic spin background induces a spin-flip in every hopping process, and each spin-flip transfers an energy $\mathcal{O}(J_{ex}) \sim v_*^2/U$ from the mobile charges to the spin sector (Golež et al. 2014). The rapid energy transfer to spins is also possible in the paramagnetic phase due to short-range spin correlations. Short-range fluctuations cannot be captured in conventional single-site DMFT, but they have been studied using cluster extensions of non-equilibrium DMFT (Eckstein and Werner 2016). The relaxation times $\tau \sim 10 - 20\, fs$ are compatible with the timescales found in the rise-time of the optical response in cuprates after a laser excitation and exact diagonalization results (Dal Conte et al. 2015).

## 5   Future Directions

As the previous examples have shown, non-equilibrium DMFT simulations have led to a number of predictions so far, mainly based on one-band model systems. By adjusting the parameter of these models, first quantitative agreement with experiment could be achieved. The status of non-equilibrium DMFT is thus comparable to equilibrium DMFT 15 years ago (Georges et al. 1996). A long-term goal is certainly to make DMFT as useful for the study of non-equilibrium states as it is nowadays for the prediction of equilibrium properties. This suggests a few developments which are subject to ongoing research:

Many phenomena in correlated materials rely on the existence of several degenerate valence orbitals. The behavior of multi-orbital systems is governed by the interplay of orbital order, spin-order, and the strong coupling to the lattice. The strong-coupling impurity solver (Eckstein and Werner 2010b) can be extended to impurity models with more than one orbital. The multi-orbital nature becomes manifest already at the shortest times, as it opens new relaxation path for photo-excited carriers (Strand et al. 2017). At long times, systems with more than one orbital may be prone to expose thermodynamically hidden states under non-equilibrium conditions when the multi-orbital nature may lead to competing orders. A controlled theoretical description of such hidden orders is an open question. A possible first step may be to measure the interactions between competing order parameters in the photo-doped state, similar to the measurement of spin-exchange interaction in the photo-doped single-band Hubbard model (Mentink and Eckstein 2014).

Furthermore, a wide range of phenomena relies on long-range interactions. Of immediate importance is the long-range Coulomb interaction, which leads to screening of the local interactions and provides a feedback of non-equilibrium distribution on the interaction parameters. In equilibrium simulations, GW+DMFT is used as a parameter-free approach to determine the interactions and the band structure, which does not suffer from the double-counting problem that arises when the ground-state DFT is combined with the finite-temperature diagrammatic DMFT approach (Biermann et al. 2003). GW+DMFT out of equilibrium provides a way to obtain the local interactions in a self-consistent way and may therefore provide the basis to a first-principle approach to correlated systems out of equilibrium (Golez et al. 2017). A full GW simulation out of equilibrium is currently not feasible because of the numerical restriction given by Kadanoff-Baym equations. However, as a first step, one can set up a multi-scale approach, in which DFT is used to obtain a band structure in a wide energy range, from which successively smaller energy windows are selected for more accurate few-band GW and DMFT+GW simulations.

# References

Amaricci A, Weber C, Capone M, Kotliar G (2012) Approach to a stationary state in a driven Hubbard model coupled to a thermostat. Phys Rev B 86:085110

Aoki H, Tsuji N, Eckstein M, Kollar M, Oka T, Werner P (2014) Nonequilibrium dynamical mean-field theory and its applications. Rev Mod Phys 86:779

Arrigoni E, Knap M, von der Linden W (2013) Nonequilibrium dynamical mean-field theory: an auxiliary quantum master equation approach. Phys Rev Lett 110:086403

Balzer K, Wolf FA, McCulloch I, Werner P, Eckstein M (2015) Nonthermal melting of Néel order in the Hubbard Model. Phys Rev X 5:031039

Bányai L, Vu QT, Mieck B, Haug H (1998) Ultrafast quantum kinetics of time-dependent RPA-screened Coulomb scattering. Phys Rev Lett 81:882

Baym G, Kadanoff LP (1961) Conservation laws and correlation functions. Phys Rev 124:287

Biermann S, Aryasetiawan F, Georges A (2003) First-principles approach to the electronic structure of strongly correlated systems: combining the GW approximation and dynamical mean-field theory. Phys Rev Lett 90:086402

Bonitz M (ed) (2000) Progress in nonequilibrium Green's functions. World Scientific, Singapore

Brandt U, Mielsch C (1991) Free energy of the Falicov-Kimball model in large dimensions. Z Phys B 82:37

Brunner H, van der Houwen PJ (1986) The numerical solution of Volterra equations. Elsevier Science Ltd ISBN-10: 0444700730

Bukov M, D'Alessio L, Polkovnikov A (2015) Universal high-frequency behavior of periodically driven systems: from dynamical stabilization to Floquet engineering. Adv Phys 64(2): 139–229

Cohen G, Gull E, Reichman D, Millis AJ (2015) Taming the dynamical sign problem in real-time evolution of quantum many-body problems. Phys Rev Lett 115:266802

Dal Conte S et al (2015) Snapshots of the retarded interaction of charge carriers with ultrafast fluctuations in cuprates. Nature Phys 11:421

Deng X, Mravlje J, Žitko R, Ferrero M, Kotliar G, Georges A (2013) How bad metals turn good: spectroscopic signatures of resilient quasiparticles. Phys Rev Lett 110:086401

Eckstein M, Kollar M (2008) Measuring correlated electron dynamics with time-resolved photoemission spectroscopy. Phys Rev B 78:245113

Eckstein M, Werner P (2010) Nonequilibrium dynamical mean-field calculations based on the noncrossing approximation and its generalizations. Phys Rev B 82:115115

Eckstein M, Werner P (2011) Thermalization of a pump-excited Mott insulator. Phys Rev B 84:035122

Eckstein M, Werner P (2013a) Nonequilibrium dynamical mean-field simulation of inhomogeneous systems. Phys Rev B 88:075135

Eckstein M, Werner P (2013b) Photo-induced states in a Mott insulator. Phys Rev Lett 110:126401

Eckstein M, Werner P (2016) Ultra-fast photo-carrier relaxation in Mott insulators with short-range spin correlations. Sci Rep 6:21235

Eckstein M, Kollar M, Werner P (2009) Thermalization after an interaction quench in the Hubbard model. Phys Rev Lett 103:056403

Eckstein M, Kollar M, Werner P (2010a) Interaction quench in the Hubbard model: relaxation of the spectral function and the optical conductivity. Phys Rev B 81:115131

Eckstein M, Oka T, Werner P (2010b) Dielectric breakdown of Mott insulators in dynamical mean-field theory. Phys Rev Lett 05:146404

Fausti D, Tobey RI, Dean N, Kaiser S, Dienst A, Hoffmann MC, Pyon S, Takayama T, Takagi H, Cavalleri A (2011) Light-induced superconductivity in a stripe-ordered cuprate. Science 331(6):189–191

Freericks JK (2008) Quenching bloch oscillations in a strongly correlated material: nonequilibrium dynamical mean-field theory. Phys Rev B 77:075109

Freericks JK, Turkowski VM, Zlatić V (2006) Nonequilibrium dynamical mean-field theory. Phys Rev Lett 97:266408

Freericks JK, Krishnamurthy HK, Pruschke T (2009) Theoretical description of time-resolved photoemission spectroscopy: application to pump-probe experiments. Phys Rev Lett 102:136401

Georges A, Kotliar G (1992) Hubbard model in infinite dimensions. Phys Rev B 45:6479

Georges A, Kotliar G, Krauth W, Rozenberg M (1996) Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. Rev Mod Phys 68:13

Giannetti C, Capone M, Fausti D, Fabrizio M, Parmigiani F, Mihailovic D (2016) Ultrafast optical spectroscopy of strongly correlated materials and high-temperature superconductors: a nonequilibrium approach. Adv Phys 65(2):58–238

Goldman N, Dalibard J (2014) Periodically driven quantum systems: effective Hamiltonians and engineered gauge fields. Phys Rev X 4(3):031027

Golež D, Bonča J, Mierzejewski M, Vidmar L (2014) Mechanism of ultrafast relaxation of a photocarrier in antiferromagnetic spin background. Phys Rev B 89:165118

Golež D, Eckstein M, Werner P (2015) Dynamics of screening in photodoped Mott insulators. Phys Rev B 92:195123

Golež D, Werner P, Eckstein M (2016) Photo-induced gap closure in an excitonic insulator. Phys Rev B 94:035121

Golez D, Boehnke L, Strand H, Eckstein M, Werner Ph (2017) Nonequilibrium GW+EDMFT: antiscreening and inverted populations from nonlocal correlations. Phys Rev Lett 118:246402

Gramsch C, Balzer K, Eckstein M, Kollar M (2013) Hamiltonian-based impurity solver for nonequilibrium dynamical mean-field theory. Phys Rev B 88:235106

Gull E, Millis AJ, Lichtenstein A, Rubtsov A, Troyer M, Werner P (2011) Continuous-time Monte Carlo methods for quantum impurity models. Rev Mod Phys 83:349

Haug H, Jauho A (2008) Quantum kinetics in transport and optics of semiconductors. Springer, Berlin

Higuchi T, Heide C, Ullmann K, Weber H, Hommelhoff P (2017) Light-field-driven currents in graphene. Nature 550:224–228

Ichikawa H et al (2011) Transient photoinduced 'hidden' phase in a manganite. Nat Mater 10(2):101–105

Imada M, Fujimori A, Tokura Y (1998) Metal insulator transitions. Rev Mod Phys 70:1039

Jotzu G, Messer M, Desbuquois R, Lebrat M, Uehlinger T, Greif D, Esslinger T (2014) Experimental realization of the topological Haldane model with ultracold fermions. Nature 515:237–240

Joura A, Freericks J, Pruschke T (2008) Steady-state nonequilibrium density of states of driven strongly correlated lattice models in infinite dimensions. Phys Rev Lett 101:196401

Kamenev A (2011) Field theory of non-equilibrium systems. Cambridge University Press, Cambridge

Kemper A, Sentef M, Moritz B, Devereaux T, Freericks JK (2017) Review of the theoretical description of time-resolved angle-resolved photoemission spectroscopy in electron-phonon mediated superconductors. Annalen der Physik 529:1600235

Knap M, Babadi M, Refael G, Martin I, Demler E (2016) Dynamical cooper pairing in nonequilibrium electron-phonon systems. Phys Rev B 94:214504

Kollar M, Wolf FA, Eckstein M (2011) Generalized Gibbs ensemble prediction of prethermalization plateaus and their relation to nonthermal steady states in integrable systems. Phys Rev B 84(5):054304

Lenarčič Z, Prelovšek P (2014) Charge recombination in undoped cuprates. Phys Rev B 90:235136

Li J, Aron C, Kotliar G, Han J (2015) Electric-field-driven resistive switching in the dissipative Hubbard model. Phys Rev Lett 114:226403

Ligges M, Avigo I, Golež D, Strand H, Stojchevska L, Kalläne M, Zhou P, Rossnagel K, Eckstein M, Werner P, Bovensiepen U (2018) Ultrafast doublon dynamics in photo-excited 1T-TaS2. Phys. Rev. Lett. 120:166401. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.166401

Lipavský P, Špička V, Velický B (1986) Generalized Kadanoff-Baym ansatz for deriving quantum transport equations. Phys Rev B 34:6933

Mahan GD (2000) Many-particle physics. Kluver Academic, New York

Mentink JH, Eckstein M (2014) Ultrafast quenching of the exchange interaction in a Mott insulator. Phys Rev Lett 113:057201

Mentink J, Balzer K, Eckstein M (2015) Ultrafast and reversible control of the exchange interaction in Mott insulators. Nature Commun 6:6708

Metzner W, Vollhardt D (1989) Correlated lattice fermions in $d = \infty$ dimensions. Phys Rev Lett 62:324

Mitrano et al (2014) Pressure-dependent relaxation in the photoexcited Mott insulator ET-$F_2$TCNQ: influence of hopping and correlations on quasiparticle recombination rates. Phys Rev Lett 112:117801

Moeckel M, Kehrein S (2008) Interaction quench in the Hubbard model. Phys Rev Lett 100(17):175702

Mühlbacher L, Rabani E (2008) Real-time path integral approach to nonequilibrium many-body quantum systems. Phys Rev Lett 100:176403

Müller-Hartmann E (1989) Correlated fermions on a lattice in high dimensions. Z Phys B 74:507

Murakami Y, Tsuji N, Eckstein M, Werner P (2017) Nonequilibrium steady states and transient dynamics of superconductors under phonon driving. Phys Rev B 96:045125

Okamoto H, Miyagoe T, Kobayashi K, Uemura H, Nishioka H, Matsuzaki H, Sawa A, Tokura Y (2010) Ultrafast charge dynamics in photoexcited $Nd_2CuO_4$ and $La_2CuO_4$ cuprate compounds investigated by femtosecond absorption spectroscopy. Phys Rev B 82:060513(R)

Onida G, Reining L, Rubio A (2002) Electronic excitations: density-functional versus many-body Green's-function approaches. Rev Mod Phys 74:601

Pavarini E, Koch E, Vollhardt D, Lichtenstein A (eds) (2014) DMFT at 25: infinite dimensions. Verlag des Forschungszentrum Jülich, Jülich

Polkovnikov A, Sengupta K, Silva A, Vengalattore M (2011) Colloquium: nonequilibrium dynamics of closed interacting quantum systems. Rev Mod Phys 83(3):863–883

Rini M, Tobey R, Dean N, Itatani J, Tomioka Y, Tokura Y, Schoenlein R, Cavalleri A (2007) Control of the electronic phase of a manganite by mode-selective vibrational excitation. Nature 449(7158):72–74

Sayyad S, Eckstein M (2016) Slow down of the electronic relaxation close to the Mott transition. Phys Rev Lett 117:096403

Schmidt P, Monien H (2002) Nonequilibrium dynamical mean-field theory of a strongly correlated system. Available via arXiv:cond-mat/0202046. Accessed 1 Jan 2017

Sciolla B, Biroli G (2013) Quantum quenches, dynamical transitions, and off-equilibrium quantum criticality. Phys Rev B 88:201110(R)

Stojchevska L, Vaskivskyi I, Mertelj T, Kusar P, Svetin D, Brazovskii S, Mihailovic D (2014) Ultrafast switching to a stable hidden quantum state in an electronic crystal. Science 344(6180):177–180

Strand H, Eckstein M, Werner P (2015) Nonequilibrium dynamical mean-field theory for bosonic lattice models. Phys Rev X 5:011038

Strand H, Golež D, Eckstein M, Werner P (2017) Hund's coupling driven photo-carrier relaxation in the two-band Mott insulator. Phys Rev B 96:165104

Strohmaier N, Greif D, Jördens R, Tarruell L, Moritz H, Esslinger T, Sensarma R, Pekker, Altman E, Demler E (2010) Observation of elastic doublon decay in the Fermi-Hubbard model. Phys Rev Lett 104:080401

Tsuji N, Oka T, Aoki H (2008) Correlated electron systems periodically driven out of equilibrium: Floquet+DMFT formalism. Phys Rev B 78:235124

Tsuji N, Oka T, Werner P, Aoki P (2011) Dynamical band flipping in fermionic lattice systems: an ac-field-driven change of the interaction from repulsive to attractive. Phys Rev Lett 106:236401

Tsuji N, Eckstein M, Werner P (2013) Nonthermal antiferromagnetic order and nonequilibrium criticality in the Hubbard model. Phys Rev Lett 110(13):136404

Werner P, Eckstein M (2013) Phonon-enhanced relaxation and excitation in the Holstein-Hubbard model Phys Rev B 88:165108

Werner P, Oka T, Millis AJ (2009) Diagrammatic Monte Carlo simulation of non-equilibrium systems. Phys Rev B 79:035320

Werner P, Tsuji N, Eckstein M (2012) Nonthermal symmetry-broken states in the strongly interacting Hubbard model. Phys Rev B 86:205101

Werner P, Held K, Eckstein M (2014) Role of impact ionization in the thermalization of photo-excited Mott insulators. Phys Rev B 90:235102

Wolf FA, McCulloch I, Schollwöck U (2014) Solving nonequilibrium dynamical mean-field theory using matrix product states. Phys Rev B 90:235131

# Correlations and Effective Interactions from First Principles Using Quantum Monte Carlo

# 19

Lucas K. Wagner

## Contents

L. K. Wagner (✉)
Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: lkwagner@illinois.edu

**Abstract**

Quantum Monte Carlo (QMC) comprises a set of techniques that use random numbers to address quantum mechanical problems. They are some of the most accurate techniques that can address large systems. This chapter gives the basics of QMC techniques on first principles models of materials.

# 1    First Principles

Any quantum system is defined by a Hilbert space $\mathscr{H}$ and Hamiltonian $\hat{H}$. Already at this level, there are several choices to be made. In all cases considered here, different choices will lead to the same results within some limit.

## 1.1    The Hilbert Space

The simplest case is for a fixed number of fermions $N$ with open boundary conditions, in which case $\mathscr{H}$ is the space of all $L^2$ normalizable antisymmetric functions from $\mathbb{R}^{3N} \to \mathbb{C}$. This Hilbert space is appropriate for an isolated molecule or other finite system. Table 1 collects a few common choices for the Hilbert space.

## 1.2    First Principles Hamiltonian

The fundamental mathematical problem is the many-particle Schrödinger equation:

$$i \frac{\partial |\Psi(t)\rangle}{\partial t} = \hat{H} |\Psi\rangle. \tag{1}$$

In first principles (ab initio) calculations of condensed matter, materials, and chemical systems, the nonrelativistic Hamiltonian is universal:

**Table 1** Hilbert spaces commonly used in first principles calculations

| | |
|---|---|
| Molecule | $\mathbb{R}^{3N} \to \mathbb{C}$, antisymmetric, $L_2$ normalizable |
| Solid (twisted boundary conditions) | $\Psi_k(r_1, r_2, \ldots, r_k + L, \ldots, r_N) = e^{i\mathbf{k}\cdot\mathbf{L}}\Psi_k(r_1, r_2, \ldots, r_k, \ldots, r_N)$, antisymmetric, normalizable on the unit cell defined by lattice vectors $\mathbf{L}$ |
| Embedding | As above, but the number of particles can vary (Fock space) |
| On a basis | Space spanned by some basis: $|\Psi\rangle = \prod c^\dagger |0\rangle$ |

$$\hat{H} = \sum_i \frac{-\hbar^2}{2m_e} \nabla_i^2 + \sum_\alpha \frac{-\hbar^2}{2m_\alpha} \nabla_\alpha^2 + \sum_{ij} \frac{e^2}{r_{ij}} + \sum_{i\alpha} \frac{-Z_\alpha e^2}{r_{i\alpha}} + \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta e^2}{r_{\alpha\beta}}, \quad (2)$$

where
$i, j$ electron indices
$\alpha, \beta$ nuclear indices
$Z_\alpha$ atomic number of nucleus $\alpha$
$m_\alpha$ mass of nucleus $\alpha$
$m_e$ mass of the electron

Throughout this chapter, we will use atomic units, in which $\hbar$, $m_e$, and $e$ are all set to 1. The unit of energy is the Hartree (27.2114 eV) and the unit of length is the Bohr (0.529 Å). This Hamiltonian is not quite perfect; it is missing relativistic effects. At the current accuracy of quantum Monte Carlo calculations, the relativistic effects are often of similar size to the solution error.

## 1.3   Born-Oppenheimer Approximation

Most electronic structure quantum Monte Carlo calculations are performed within the Born-Oppenheimer approximation, in which the *electronic* Hamiltonian is used. It is not absolutely necessary to do this; QMC techniques can operate with fully quantum nuclei and electrons. However, the approximation is often quite small and improves the efficiency and simplicity of the algorithms by a large amount.

## 1.4   Effective Core Potentials

For nuclei with large $Z$, the core electrons increase the variance of the total energy, which decreases the efficiency of Monte Carlo. However, these electrons do not affect the physics and chemistry of the material very much at the energies of interest. For that reason, it is often a good trade-off in terms of accuracy/efficiency to replace the core electrons with an effective core potential:

$$\frac{Z_\alpha}{r_{\alpha i}} \rightarrow \frac{Z_{eff,\alpha}}{r_{i\alpha}} + V_l(r_{i\alpha}) + \sum_{\ell m} V_\ell(r_{i\alpha}) |lm\rangle \langle lm| . \quad (3)$$

This must be done with high accuracy; if density functional theory (DFT) potentials are used, then the QMC calculation will be limited mainly by the accuracy of the potential. There are now several sets of effective core potentials (Trail and Needs 2013; Burkatzki et al. 2007, 2008) meant for QMC calculations that retain accuracy and are smooth for efficiency purposes. These are chosen so that the ions (nucleus plus the core electrons) emulate accurate calculations of the full system. They also

typically include at least averaged relativistic effects and may include polarization terms and non-scalar relativistic effects like $\mathbf{L} \cdot \mathbf{S}$ coupling.

## 1.5   Useful Mathematical Relationships

Some of these relationships are available in most books on quantum mechanics; we remind the reader of them here and clarify notation.

$\hat{H}$ can be diagonalized. There is a set of states $\{|\Phi_i\rangle\}$ such that

$$\hat{H} = \sum_i |\Phi_i\rangle\, E_i\, \langle\Phi_i| . \tag{4}$$

The state with lowest energy $E_0$ is called the ground state $|\Phi_0\rangle$. If $|\Psi\rangle \in \mathcal{H}$, then it has a spectral expansion $|\Psi\rangle = \sum_i c_i |\Phi_i\rangle$, for some complex numbers $c_i$. If $|\Psi\rangle \in \mathcal{H}$, then by the variational principle:

$$\langle\Psi|\hat{H}|\Psi\rangle \geq E_0. \tag{5}$$

This is provable using the above relationships.

**Projection**  Suppose that $|\Psi(\tau)\rangle = \exp\left(-\tau\hat{H}\right)|\Psi(0)\rangle$. Then $|\Psi(\tau)\rangle = \sum_i c_i \exp(-\tau E_i)|\Phi_i\rangle$ and

$$\frac{|\Psi(\tau)\rangle}{\sqrt{\langle\Psi(\tau)|\Psi(\tau)\rangle}} = \sum_i \frac{c_i \exp(-\tau E_i)}{\sqrt{\sum_j c_j^2 \exp\left(-2\tau E_j\right)}} |\Phi_i\rangle . \tag{6}$$

Define

$$c_i(\tau) = \frac{c_i \exp(-\tau E_i)}{\sqrt{\sum_j c_j^2 \exp\left(-2\tau E_j\right)}}. \tag{7}$$

Suppose that there is a $k$ such that $E_k < E_i$. Then we can divide the numerator and denominator by $\exp(-\tau E_k)$:

$$c_i(\tau) = \frac{c_i \exp(-\tau(E_i - E_k))}{\sqrt{c_k^2 + \sum_{j \neq k} c_j^2 \exp\left(-2\tau(E_j - E_k)\right)}}. \tag{8}$$

So long as $c_k \neq 0$, $c_i(\tau)$ goes to zero exponentially as $\tau \to \infty$. The only remaining nonzero coefficient is the lowest eigenstate with a nonzero component of $|\Psi\rangle$. This is an example of the power method, in which an operator is applied repeatedly to obtain its largest eigenstate.

**Mixed estimator for energy**   As long as $\langle \Phi_i | \Psi \rangle \neq 0$,

$$\frac{\langle \Phi_i | \hat{H} | \Psi \rangle}{\langle \Phi_i | \Psi \rangle} = E_i. \tag{9}$$

This can be derived by using $\hat{H} | \Phi_i \rangle = E_i | \Phi_i \rangle$.

As a result of Eqs. 8 and 9, if $\langle \Psi | \Phi_0 \rangle \neq 0$,

$$\frac{\langle \Psi | \exp\left(-\tau \hat{H}\right) \hat{H} | \Psi \rangle}{\langle \Psi | \exp\left(-\tau \hat{H}\right) | \Psi \rangle} = E_0, \tag{10}$$

the ground state energy.

## 2   Monte Carlo

The main use of Monte Carlo in this chapter will be to evaluate many-dimensional integrals. Suppose that $\mathbf{X}$ is some vector. Then

$$\int f(\mathbf{X}) \rho(\mathbf{X}) d\mathbf{X} = \langle f(\mathbf{X}) \rangle_{\mathbf{X} \sim \rho}. \tag{11}$$

Here, $\mathbf{X} \sim \rho$ means that $\mathbf{X}$ is sampled from the probability density $\rho$, and the angle brackets $\langle \rangle$ denote the average. For this to be true, $\rho(\mathbf{X})$ must be a probability density ($\rho(\mathbf{X}) \geq 0$ and $\int \rho(\mathbf{X}) dX = 1$), but $f(\mathbf{X})$ may be any function.

Monte Carlo evaluation of an integral consists of two parts: sampling and averaging. Sampling means to generate $\mathbf{X}$ with probability proportional to $\rho(\mathbf{X})$. Averaging means evaluation of $f(\mathbf{X})$ and evaluation of statistical uncertainties.

Suppose that we generate some number $N$ of $\mathbf{X}_i \sim \rho$. For the moment, we will assume that the $\mathbf{X}_i$ are independent, although this is not necessary. Then we would like to know the distribution of the average variable $f_N = \frac{1}{N} \sum_i^N f(\mathbf{X}_i)$. Let the variance of $f$ be

$$\mathrm{Var}_f = \left\langle (f(\mathbf{X}) - \langle f(\mathbf{X}) \rangle)^2 \right\rangle_{\mathbf{X} \sim \rho}. \tag{12}$$

If $\mathrm{Var}_f$ is finite and in the limit of large $N$, then the central limit theorem states that $f_N$ has a normal distribution centered around the exact average and with reduced variance:

$$f_N \sim \mathcal{N}\left(\langle f \rangle, \frac{\mathrm{Var}_f}{N}\right). \tag{13}$$

Every Monte Carlo calculation should include an estimate of the variance of the average. On plots, the square root of the variance (the standard deviation) is typically represented as an "error bar."

## 3    Trial Functions

Trial functions are many-particle wave functions. Typically there are some parameters $\mathbf{P}$ in the wave function. $\mathbf{x}$ is the position ($\mathbf{r}$) and spin ($\omega$) coordinate of an electron. $\chi_i$ is a function of $\mathbf{x}$ which produces a complex number. $N$ is the number of electrons. Much of the code in quantum Monte Carlo programs is devoted to evaluating trial wave functions efficiently.

Quantum Monte Carlo brings two major improvements over deterministic methods based on variational wave functions. First, a complex Jastrow factor can be used, which improves the efficiency and accuracy dramatically. While a single Slater determinant is often not a good approximation to an interacting electronic system, a Slater determinant multiplied by a Jastrow factor is much better. Second, projection Monte Carlo techniques can be used which further improve the accuracy beyond an analytic form.

## 3.1    Slater Determinant

Then

$$\Psi_S(\mathbf{x}_1, \mathbf{x}_2, \ldots; \chi_1, \chi_2, \ldots, \chi_N) = \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \chi_1(\mathbf{r}_3) & \cdots & \chi_1(\mathbf{x}_N) \\ \chi_2(\mathbf{x}_1) & \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{r}_3) & \cdots & \chi_2(\mathbf{x}_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \chi_N(\mathbf{x}_1) & \chi_N(\mathbf{x}_2) & \chi_N(\mathbf{r}_3) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix}$$

is antisymmetric on exchange of spin and position. It is a functional of the spin orbitals.

For spin collinear calculations, one typically works in a basis of $S_z$ eigenstates, so there is a constant $N_\uparrow$ and $N_\downarrow$.

$$\chi_i(\mathbf{x}) = \begin{cases} \phi_i(\mathbf{r})\alpha(\omega) & \text{if } i \leq N_\uparrow \\ \phi_i(\mathbf{r})\beta(\omega) & \text{if } i > N_\uparrow \end{cases} \tag{14}$$

Assume $N_\uparrow \geq N_\downarrow$. Then the real-space orbitals $\phi_i$ can be either restricted, which means that $\phi_i = \phi_{i+N_\uparrow}$ for $i = 1 \ldots N_\uparrow$, or they can be unrestricted, which relaxes this condition.

**Table 2** Common basis functions used in quantum Monte Carlo

| Name | Basis type |
|------|-----------|
| Blip | Localized finite support basis |
| Gaussian | Sum of Gaussian basis functions centered around atoms |
| Slater | Sum of exponential basis functions centered around atoms |
| Localized numerical | Numerical basis functions centered around atoms |
| Plane waves | Periodic plane wave functions |

A given real-space orbital $\phi_i$ is parameterized in a basis $\{b_j\}$:

$$\phi_i(\mathbf{r}) = \sum_j c_{ij} b_j(\mathbf{r}). \tag{15}$$

The trial function is thus specified by the basis functions $\{b_j\}$ (see Table 2) and the coefficient matrix $c_{ij}$. These may be optimized directly within the variational Monte Carlo framework (Toulouse and Umrigar 2008) or taken from a density functional theory or quantum chemistry program. Most QMC packages provide conversion tools for this operation. In some cases, a hybrid approach is used (Wagner and Mitas 2003), in which the orbitals are generated using different density functionals.

It is also possible to perform noncollinear calculations using quantum Monte Carlo (Melton et al. 2016). In that case, $\omega$ is parameterized and allowed to dynamically change.

## 3.2 Jastrow Factor

The Jastrow factor (Bijl 1940; Dingle 1949; Jastrow 1955) is an all-positive factor that includes electron-electron correlations:

$$\Psi_J(\mathbf{x}_1, \mathbf{x}_2, \ldots; \mathbf{P}) = \exp\left[\underbrace{\sum_{ij} \chi(\mathbf{x}_i, \mathbf{x}_j; \mathbf{P}_{\text{cusp}})}_{\text{Cusp condition}} + \underbrace{\sum_{ij} f(\mathbf{r}_i, \mathbf{r}_j; \mathbf{P}_{\text{Correlation}})}_{\text{Correlation factor}}\right] \tag{16}$$

For electronic systems, a high-quality Jastrow factor can account for between 50% and 85% of the correlation energy with a few parameters. It is related to the $f_{12}$ and transcorrelated methods used in quantum chemistry.

The cusp part of the Jastrow factor enforces the Kato cusp conditions. The correlation part can be expanded in many ways, which can differ from implementation to implementation. It appears that different parameterizations obtain quite similar results (Doblhoff-Dier et al. 2016).

### 3.3 Multiple Slater Determinants

As in quantum chemistry calculations, the many-body wave function can be expanded as a sum of Slater determinants. When multiplied by a Jastrow factor, the convergence in determinants is much faster than without (Umrigar et al. 2007). Determinants can be selected from a configuration interaction calculation, one of various selected configuration interaction calculations, or directly within quantum Monte Carlo (Holzmann et al. 2003; Williams and Wagner 2016).

### 3.4 Backflow and Iterated Backflow

Construct a dressed coordinate:

$$\mathbf{q}_i = \mathbf{r}_i + \sum_{j \neq i} f(\mathbf{r}_j, \mathbf{r}_i), \tag{17}$$

where $f$ is a vector-valued function. Then instead of $\mathbf{x}_i = (\mathbf{r}_i, \omega)$, use $\mathbf{x}_i = (\mathbf{q}_i, \omega)$ in the Slater determinant (or pairing wave function). There are several ways to achieve this (Kwon et al. 1993; López Ríos et al. 2006; Taddei et al. 2015). As in the original application (Feynman and Cohen 1956), this wave function is particularly effective for metallic systems. If the function $f$ is efficient to evaluate, then the Slater determinant can be evaluated with a moderate increase in computational time (Filippi et al. 2016).

### 3.5 Tensor Networks, Correlator, and Matrix Product States

The broad class of tensor network states (Wouters et al. 2014) and correlator product states (Changlani et al. 2009) have been used in some quantum Monte Carlo studies as trial wave functions. These work essentially and similarly to a determinant expansion, since they are written in a second-quantized notation. These functions have not been used very much in first principles calculations and so this chapter will not focus on them.

### 3.6 Pairing Wave Functions: Geminal and Pfaffians

The most general form of the pairing wave function is the Pfaffian (Bajdich et al. 2006):

$$\mathrm{Pf} \begin{bmatrix} \xi^{\uparrow\uparrow} & \Phi^{\uparrow\downarrow} & \phi^{\uparrow} \\ -\Phi^{\uparrow\downarrow T} & \xi^{\downarrow\downarrow} & \phi^{\uparrow} \\ -\phi^{\uparrow T} & -\phi^{\uparrow T} & 0 \end{bmatrix}, \tag{18}$$

where each of the Greek letters in the matrix is a sub-matrix that correlates electrons of the corresponding spins and the $\phi$ sub-matrices address unpaired electrons. This function is antisymmetric due to the properties of the Pfaffian, which can be computed in a similar amount of time as a determinant.

If the same-spin electrons are not correlated ($\xi = 0$ in Eq. 18), Eq. 18 reduces to the antisymmetrized geminal product, which only correlates the different-spin electrons. This is the same as the number-projected Bardeen-Cooper-Schrieffer wave function. In practice, these wave functions suffer from severe size-consistency issues which require them to be used with a Jastrow factor (Neuscamman 2012).

## 4  Quantum Monte Carlo Methods

### 4.1  Variational Monte Carlo

Parameterize a wave function according to some parameters $\mathbf{P}$. Then the expectation value of the energy is

$$E(\mathbf{P}) = \frac{\langle \Psi(\mathbf{P}) | \hat{H} | \Psi(\mathbf{P}) \rangle}{\langle \Psi(\mathbf{P}) | \Psi(\mathbf{P}) \rangle}. \tag{19}$$

This expectation value can be written as an integral in some basis $|b\rangle$ (suppressing the $\mathbf{P}$ argument from $\Psi$):

$$E(\mathbf{P}) = \frac{\int \langle \Psi | b \rangle \langle b | H | \Psi \rangle db}{\int \langle \Psi | b \rangle \langle b | \Psi \rangle db} \tag{20}$$

$$= \int \underbrace{\frac{\langle \Psi | b \rangle \langle b | \Psi \rangle}{\int \langle \Psi | b \rangle \langle b | \Psi \rangle db}}_{\rho(b)} \underbrace{\frac{\langle b | H | \Psi \rangle}{\langle b | \Psi \rangle}}_{f(b)} db, \tag{21}$$

where $\rho(b)$ and $f(b)$ refer to Eq. 11. It is most common for the basis $b$ to be the position, but other basis sets are possible (Neuscamman 2012).

Using the variational principle (Eq. 5), $E(\mathbf{P}) \geq E_0$. So minimizing $E(\mathbf{P})$ gives an upper bound to the ground state energy. The current state of the art uses the linear method (Umrigar et al. 2007) to accomplish this. In this method, the wave function is approximated using a first-order Taylor expansion:

$$|\Psi(\mathbf{P})\rangle = |\Psi(\mathbf{P}_0)\rangle + \sum_i \delta p_i \left. \frac{\partial |\Psi(\mathbf{P})\rangle}{\partial p_i} \right|_{\mathbf{P}=\mathbf{P}_0}. \tag{22}$$

Minimizing $E(\mathbf{P})$ results in a generalized eigenvalue problem. The reason this method is efficient is that it is possible to evaluate the necessary matrix elements

with very low variance. A major theme of advances in variational Monte Carlo techniques is the reduction of variance while maintaining zero bias.

## 4.2    Projection Monte Carlo

Projection Monte Carlo techniques are stochastic implementations of the power method. Let the projection operator with parameter $\tau$ be $\hat{P}(\tau)$. Typically, as $\tau \to 0$, $\hat{P}$ goes to the identity operator, and often there is an accuracy/efficiency trade-off as $\tau$ is increased. Formally, we wish to stochastically implement the operation:

$$|\Psi(N\tau)\rangle = \left(\hat{P}(\tau)\right)^{N} |\Psi(0)\rangle . \tag{23}$$

As $N \to \infty$, $|\Psi(N\tau)\rangle$ will be the desired state, as shown in Eq. 8 for the $e^{-\tau\hat{H}}$ operator.

The stochastic implementation of projection can be understood as a Monte Carlo integration. Insert identity operators in a many-body basis $b$ in Eq. 23:

$$\langle b_0|\Psi(N\tau)\rangle = \int \left[\prod_{i=1}^{N}\langle b_{i-1}|\hat{P}(\tau)|b_i\rangle\right] \langle b_N|\Psi(0)\rangle db_1 \dots db_N. \tag{24}$$

The set $\{b_i\}$ of $N$ many-body coordinates in the basis space is called the *path*. It must be the case that the matrix element $\langle b_{i-1}|\hat{P}(\tau)|b_i\rangle$ is known. This is often why $\tau$ must be relatively small; for the example of $\hat{P}(\tau) = \exp\left(-\tau\hat{H}\right)$, the Trotter-Suzuki (Trotter 1959; Suzuki 1976) expansion can be used for small enough $\tau$. To simplify the notation, set

$$\langle b_{i-1}|\hat{P}(\tau)|b_i\rangle = P_{i-1,i} \tag{25}$$

and

$$B = \{b_i\} \tag{26}$$

To fully translate Eq. 24 into a Monte Carlo integral, we must identify a probability density $\rho$ to sample and a function $f$ to evaluate. The main constraints are that $\rho$ must be a probability density and that the variance of $f$ must be finite and small enough so that its expectation value can be determined using the central limit theorem efficiently. There are many choices of these functions, which can affect efficiency and biases of the final algorithm. In this chapter, we will focus on the *mixed estimator* (Eq. 10), which allows for efficient evaluation of the ground state energy. This estimator is

$$\frac{\langle \Psi(0)|\hat{H}|\Psi(N\tau)\rangle}{\langle \Psi(0)|\Psi(N\tau)\rangle} = E_0 = \qquad (27)$$

$$\int \underbrace{\frac{\langle \Psi(0)|\hat{H}|b_0\rangle}{\langle \Psi(0)|b_0\rangle}}_{\tilde{f}} \underbrace{\frac{\langle \Psi(0)|b_0\rangle \left[\prod_{i=1}^{N} P_{i-1,i}\right] \langle b_N|\Psi(0)\rangle}{\int \langle \Psi(0)|b_0\rangle \left[\prod_{i=1}^{N} P_{i-1,i}\right] \langle b_N|\Psi(0)\rangle db_1 \ldots db_N}}_{\tilde{\rho}} \, db_1 \ldots db_N$$

$$(28)$$

$\tilde{\rho}(B)$ is not quite a probability density yet, since it may be negative or in some cases it can be complex. In some cases, including bosonic systems and a number of model Hamiltonians, a formulation of $\tilde{\rho}$ and $\tilde{f}$ can be found such that $\tilde{\rho}$ is positive and the variance of $\tilde{f}$ is small enough so that the method can be efficient. In other techniques, such as the released node or transient, the signs of $\tilde{\rho}$ are put into $\tilde{f}$, which increases the variance of $\tilde{f}$ dramatically. This variance increases exponentially with system size, which is called the *sign problem.*

In cases where the sign problem cannot be removed, one is either left with an algorithm that scales exponentially, or an approximation must be made. The simplest approach is to approximate $\tilde{\rho}$ such that it is always positive. For example, one can take the following weight function:

$$w(B) = \begin{cases} \tilde{\rho}(B) & \text{if } \tilde{\rho}(B) \geq 0 \\ 0 & \text{if } \tilde{\rho}(B) < 0 \end{cases}. \qquad (29)$$

Then one can sample a probability density $\rho(B) = \frac{w(B)}{\int w(B)dB}$. This can be done in several ways, as noted in Sect. 5.3.

The differences between projector Monte Carlo methods consist of what approximation (if any) is made for $\rho$ and $f$ and what basis $B$ is used to express the path and matrix elements. Some of the methods write their basis in second-quantized form, while others write them in first-quantized form. For a detailed discussion, see Kolodrubetz et al. (2013) and Umrigar (2015). These choices interact with one another, and there are many variations and trade-offs between them. Table 3 lists several common projector Monte Carlo methods and their choices.

## 5 Simulation Errors

### 5.1 Time Step Errors

Suppose that we use the projector $\exp\left(-\tau \hat{H}\right)$ and are working in the real-space basis. Then we wish to evaluate the matrix elements:

**Table 3** Projection operators and basis used in different Monte Carlo techniques

| Method | Basis | Quantization | Operator | Stabilization |
|---|---|---|---|---|
| DMC | **R** | First | $\exp\left(-\tau\hat{H}\right)$ | Fixed node/phase |
| Reptation Monte Carlo[a] | **R** | First | $\exp\left(-\tau\hat{H}\right)$ | Fixed node/phase |
| Full CI QMC | Finite basis | Second | $(1-\tau\hat{H})$ | Initiator |
| Auxiliary field QMC | Finite basis | Second | $\exp\left(-\tau\hat{H}\right)$ | Phaseless |
| Lattice regularized DMC | Lattice on **R** | First | $(1-\tau\hat{H})$ | Fixed node/phase |
| Green function MC | **R** | First | $(1-\tau\hat{H})^{-1}$ | Fixed node/phase |

[a]Also known as variational path integrals and path integrals for the ground state

$$\langle\mathbf{R}|\exp\left(-\tau\hat{H}\right)|\mathbf{R}'\rangle. \tag{30}$$

This is typically done in the small $\tau$ approximation and is a rather lengthy subject. A detailed discussion for this operator on electronic systems can be found in Umrigar et al. (1993).

## 5.2    Finite Size Effects

Bulk systems are typically approximated using twisted boundary conditions, in which the bulk wave function is approximated as wave function with the following constraint:

$$\Psi_{\mathbf{k}}(\mathbf{r}_1, \mathbf{r}_2 + \mathbf{L}, \ldots, \mathbf{r}_N) = \exp(i\mathbf{k}\cdot\mathbf{L})\Psi_{\mathbf{k}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N), \tag{31}$$

where $\mathbf{L}$ is a lattice vector of the simulation cell. The infinite limit is taken by increasing the size of the simulation cell and the number of particles, keeping the density $N/V$ constant.

There are two main sources of finite size errors, corresponding to the kinetic and potential energy of the Hamiltonian. Kinetic energy errors can be corrected by considering many values of $\mathbf{k}$, so-called twist averaging. Errors due to the Coulomb interaction can be removed in several ways, either through embedding (Drummond et al. 2008) or using the electron structure factor (Chiesa et al. 2006).

## 5.3    Fixed Node/Sign Control Errors

In many cases, projection Monte Carlo methods have a sign problem as discussed in the relevant section. This sign problem makes exact calculations untenable except for quite small systems. Different techniques approach this sign error in different ways, noted in Table 3.
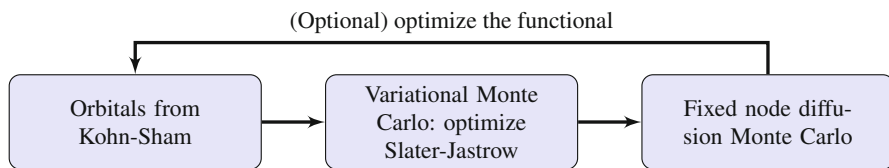
**Fig. 1** Workflow for FN-DMC calculations of materials

## 6 Variational + Fixed Node Diffusion Monte Carlo

This chapter has so far focused on general statements about quantum Monte Carlo techniques for ab initio systems. This section will focus on a standard operating procedure that has emerged in the past decades. It offers a good trade-off between accuracy, efficiency, and complexity of the calculation (Fig. 1).

### 6.1 Hamiltonian

Since core electrons are very expensive to simulate, they are typically replaced with an effective core potential. Potentials based on density functional theory are typically not accurate enough (Saritas et al. 2017), so correlation-consistent potentials must be used (Burkatzki et al. 2007, 2008; Trail and Needs 2013; Bennett et al. 2017). Modern algorithms use the T-move (Casula 2006; Casula et al. 2010) method to handle the nonlocal operator in the projector.

### 6.2 Trial Wave Function

The trial wave function $|\Psi_T\rangle$ is the Slater-Jastrow wave function, which consists of a Slater wave function multiplied by a Jastrow correlation factor. The orbitals in the Slater determinant are typically taken from Kohn-Sham density functional theory calculations. The functional used can affect the quality of the orbitals; often using hybrid DFT orbitals offer good quality without needing to optimize the orbitals within QMC. The Jastrow factor is typically energy optimized (Umrigar et al. 2007), which leads to better wave functions than variance optimizations.

### 6.3 Fixed Node Diffusion Monte Carlo (FN-DMC)

Application of FN-DMC to the trial wave function results in an energy dependent on several factors: $E_{DMC}(\{\phi_i\}, \tau, N)$, where $\{\phi_i\}$ are the orbitals, $\tau$ is the time step, and $N$ is the number of electrons in the simulation. If $\tau$ is small enough and $N$ is large enough, then

$$E_{DMC}(\{\phi_i\}, \tau, N) = E_{DMC}(\{\phi_i\}, 0, N) + \tau \varepsilon_{\text{time step}}(\{\phi_i\}, N). \tag{32}$$

One thus performs the calculation for several values of $\tau$ and extrapolates the result to zero to obtain $E_{DMC}(\{\phi_i\}, 0, N)$. Then

$$E_{DMC}(\{\phi_i\}, 0, N) = E_{DMC}(\{\phi_i\}, 0, \infty) + \varepsilon_{\text{finite size}}(\{\phi_i\})/N. \tag{33}$$

Finally, $E_{DMC}(\{\phi_i\}, 0, \infty)$ is an upper bound to the exact ground state energy. One can vary $\{\phi_i\}$ to obtain the lowest upper bound.

## 7    Effective Interactions and Models from Quantum Monte Carlo

The stabilization method for projector Monte Carlo restricts what wave function can be obtained using the technique. One way to understand this is that the final wave function after projection is now a *functional* of the guiding wave function: $|\Phi_{\text{projected}}[\Psi_G]\rangle$. For brevity, we will just call this $|\Phi[g]\rangle$, where $g$ is the guiding wave function. Because of the projection operation, $|\Phi[g]\rangle$ contains only small components of high-energy wave functions. By varying $g$, one can obtain a sampling of wave functions with relatively low energy compared to the overall spectrum. As noted in Sect. 6, $g$ can be varied to obtain minimum upper bounds, if a variational way of evaluating the energy expectation value can be found. However, these wave functions can also be used to explore effective models. The detailed theory is available in a recent publication by Changlani et al. (2015).

Consider the objective of estimating an electronic gap using quantum Monte Carlo. The effective model of the material is $\sum_{i\sigma} \varepsilon_i \hat{n}_{i\sigma}$, where $\hat{n}_i$ is the orbital occupation operator. Then

$$E[g] = \sum_{i\sigma} \varepsilon_i \langle \Psi[g] | \hat{n}_{i\sigma} | \Psi[g] \rangle, \tag{34}$$

where $\Psi[g]$ is only allowed to vary along the lowest-energy wave functions that change the orbital occupation. From this equation, one can compute $\varepsilon_i - \varepsilon_j$ by varying the occupation. This is typically done in solids by using two trial functions: Slater-Jastrow wave functions with different occupations of KS orbitals. Empirically, this procedure obtains results that are in close agreement to experimental values, with around 0.1 eV errors.

The same procedure can be used to estimate superexchange constants using QMC. In this case the effective model is

$$E[g] = \sum_{ij} J_{ij} \langle \Psi[g] | S_i \cdot S_j | \Psi[g] \rangle, \tag{35}$$

so one should sample projected wave functions that differ in their spin orientations.

An example of how this works for a spin dimer. It is relatively easy to compute the frozen spin states of the dimer, $|\uparrow\uparrow\rangle$ and $|\uparrow\downarrow\rangle$. For a dimer Heisenberg model,

$$H = J S_1 \cdot S_2 = \frac{1}{2} J \left[ (S_1 + S_2)^2 - S_2^2 - S_1^2 \right] = \frac{1}{2} J \left[ s(s+1) - 3/2 \right] . \tag{36}$$

For the two frozen spin states:

$$E(\uparrow\uparrow) = \langle \uparrow\uparrow | J S_1 \cdot S_2 | \uparrow\uparrow \rangle = J/4 \tag{37}$$

and

$$E(\uparrow\downarrow) = \langle \uparrow\downarrow | J S_1 \cdot S_2 | \uparrow\downarrow \rangle = -J/4. \tag{38}$$

The energy difference between these two states is thus $J/2$. The singlet eigenstate $\sqrt{\frac{1}{2}}\left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\right)$ (with $s = 0$) has energy $-3J/4$, while the triplet eigenstate $|\uparrow\uparrow\rangle$ (with $s = 1$) has energy $J/4$, so the singlet-triplet excitation energy is $J$.

## 8    Conclusions and outlook

At the time of this writing (2018), quantum Monte Carlo techniques have been applied to many materials systems. A semi-complete list of strongly correlated systems studied using these techniques was compiled in 2016 (Wagner and Ceperley 2016). Excitation energy gaps, equations of state, including bulk moduli and lattice constants, and ground state magnetic orderings are now a matter of routine for these techniques with modern program packages. Nonenergetic ground state properties like electron density, spin density, and static correlation functions are also computed routinely. A very rough limit on the size of systems treatable with modern computational resources is around 1000 valence electrons, although this number can vary dramatically depending on the energy scale and variance of the quantity of interest. Currently, this is very much a moving target and so we will not go very much into detail; we refer readers to ▶ Chap. 46, "Quantum Monte Carlo for Electronic Systems Containing $d$ and $f$ Electrons" in this volume for more details.

## References

Bajdich M, Mitas L, Drobný G, Wagner LK, Schmidt KE (2006) Pfaffian pairing wave functions in electronic-structure quantum Monte Carlo simulations. Phys Rev Lett 96(13):130201. http://link.aps.org/doi/10.1103/PhysRevLett.96.130201

Bennett MC, Melton CA, Annaberdiyev A, Wang G, Shulenburger L, Mitas L (2017) A new generation of effective core potentials for correlated calculations. J Chem Phys 147(22):224106. http://aip.scitation.org/doi/abs/10.1063/1.4995643

Bijl A (1940) The lowest wave function of the symmetrical many particles system. Physica 7(9):869–886. https://doi.org/10.1016/0031-8914(40)90166-5, http://www.sciencedirect.com/science/article/pii/0031891440901665

Burkatzki M, Filippi C, Dolg M (2007) Energy-consistent pseudopotentials for quantum Monte Carlo calculations. J Chem Phys 126(23):234105. https://doi.org/10.1063/1.2741534

Burkatzki M, Filippi C, Dolg M (2008) Energy-consistent small-core pseudopotentials for 3D-transition metals adapted to quantum Monte Carlo calculations. J Chem Phys 129(16):164115. https://doi.org/10.1063/1.2987872

Casula M (2006) Beyond the locality approximation in the standard diffusion Monte Carlo method. Phys Rev B 74(16):161102(R). http://link.aps.org/doi/10.1103/PhysRevB.74.161102

Casula M, Moroni S, Sorella S, Filippi C (2010) Size-consistent variational approaches to nonlocal pseudopotentials: standard and lattice regularized diffusion Monte Carlo methods revisited. J Chem Phys 132(15):154113. http://scitation.aip.org/content/aip/journal/jcp/132/15/10.1063/1.3380831

Changlani HJ, Kinder JM, Umrigar CJ, Chan GKL (2009) Approximating strongly correlated wave functions with correlator product states. Phys Rev B 80(24):245116. http://link.aps.org/doi/10.1103/PhysRevB.80.245116

Changlani HJ, Zheng H, Wagner LK (2015) Density-matrix based determination of low-energy model Hamiltonians from ab initio wavefunctions. J Chem Phys 143(10):102814. http://scitation.aip.org/content/aip/journal/jcp/143/10/10.1063/1.4927664

Chiesa S, Ceperley D, Martin R, Holzmann M (2006) Finite-size error in many-body simulations with long-range interactions. Phys Rev Lett 97(7):076404. http://link.aps.org/doi/10.1103/PhysRevLett.97.076404

Dingle RB (1949) LI. The zero-point energy of a system of particles. Lond Edinb Dublin Philos Mag J Sci 40(304):573–578. https://doi.org/10.1080/14786444908521743

Doblhoff-Dier K, Meyer J, Hoggan PE, Kroes GJ, Wagner LK (2016) Diffusion Monte Carlo for accurate dissociation energies of 3D transition metal containing molecules. J Chem Theory Comput 12(6):2583–2597. http://pubs.acs.org/doi/abs/10.1021/acs.jctc.6b00160

Drummond N, Needs R, Sorouri A, Foulkes W (2008) Finite-size errors in continuum quantum Monte Carlo calculations. Phys Rev B 78(12):125106. http://link.aps.org/doi/10.1103/PhysRevB.78.125106

Feynman RP, Cohen M (1956) Energy spectrum of the excitations in liquid helium. Phys Rev 102(5):1189–1204. http://link.aps.org/doi/10.1103/PhysRev.102.1189

Filippi C, Assaraf R, Moroni S (2016) Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. J Chem Phys 144(19):194105. http://scitation.aip.org/content/aip/journal/jcp/144/19/10.1063/1.4948778

Holzmann M, Ceperley DM, Pierleoni C, Esler K (2003) Backflow correlations for the electron gas and metallic hydrogen. Phys Rev E 68(4):046707. http://link.aps.org/doi/10.1103/PhysRevE.68.046707

Jastrow R (1955) Many-body problem with strong forces. Phys Rev 98(5):1479–1484. https://link.aps.org/doi/10.1103/PhysRev.98.1479

Kolodrubetz MH, Spencer JS, Clark BK, Foulkes WM (2013) The effect of quantization on the full configuration interaction quantum Monte Carlo sign problem. J Chem Phys 138(2):024110. http://aip.scitation.org/doi/abs/10.1063/1.4773819

Kwon Y, Ceperley DM, Martin RM (1993) Effects of three-body and backflow correlations in the two-dimensional electron gas. Phys Rev B 48(16):12037. http://link.aps.org.libproxy.mit.edu/doi/10.1103/PhysRevB.48.12037

López Ríos P, Ma A, Drummond ND, Towler MD, Needs RJ (2006) Inhomogeneous backflow transformations in quantum Monte Carlo calculations. Phys Rev E 74(6):066701. http://link.aps.org/doi/10.1103/PhysRevE.74.066701

Melton CA, Zhu M, Guo S, Ambrosetti A, Pederiva F, Mitas L (2016) Spin-orbit interactions in electronic structure quantum Monte Carlo methods. Phys Rev A 93(4):042502. http://link.aps.org/doi/10.1103/PhysRevA.93.042502

Neuscamman E (2012) Size consistency error in the antisymmetric geminal power wave function can be completely removed. Phys Rev Lett 109(20):203001. http://link.aps.org/doi/10.1103/PhysRevLett.109.203001

Saritas K, Mueller T, Wagner L, Grossman JC (2017) Investigation of a quantum Monte Carlo protocol to achieve high accuracy and high-throughput materials formation energies. J Chem Theory Comput 13(5):1943–1951. https://doi.org/10.1021/acs.jctc.6b01179

Suzuki M (1976) Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems. Commun Math Phys 51(2):183–190. https://link.springer.com/article/10.1007/BF01609348

Taddei M, Ruggeri M, Moroni S, Holzmann M (2015) Iterative backflow renormalization procedure for many-body ground-state wave functions of strongly interacting normal Fermi liquids. Phys Rev B 91(11):115106. https://link.aps.org/doi/10.1103/PhysRevB.91.115106

Toulouse J, Umrigar CJ (2008) Full optimization of Jastrow–Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules. J Chem Phys 128(17):174101

Trail JR, Needs RJ (2013) Pseudopotentials for correlated electron systems. J Chem Phys 139(1):014101. http://aip.scitation.org/doi/10.1063/1.4811651

Trotter HF (1959) On the product of semi-groups of operators. Proc Am Math Soc 10(4):545–551. http://www.jstor.org/stable/2033649

Umrigar CJ (2015) Observations on variational and projector Monte Carlo methods. J Chem Phys 143(16):164105

Umrigar CJ, Nightingale MP, Runge KJ (1993) A diffusion Monte Carlo algorithm with very small time-step errors. J Chem Phys 99(4):2865–2890. https://doi.org/10.1063/1.465195

Umrigar CJ, Toulouse J, Filippi C, Sorella S, Hennig RG (2007) Alleviation of the Fermion-Sign problem by optimization of many-body wave functions. Phys Rev Lett 98(11):110201. http://link.aps.org/doi/10.1103/PhysRevLett.98.110201

Wagner L, Mitas L (2003) A quantum Monte Carlo study of electron correlation in transition metal oxygen molecules. Chem Phys Lett 370(3–4):412–417. https://doi.org/10.1016/S0009-2614(03)00128-3, http://www.sciencedirect.com/science/article/pii/S0009261403001283

Williams KT, Wagner LK (2016) Using local operator fluctuations to identify wave function improvements. Phys Rev E 94(1):013303. http://link.aps.org/doi/10.1103/PhysRevE.94.013303

Wouters S, Verstichel B, Van Neck D, Chan GKL (2014) Projector quantum Monte Carlo with matrix product states. Phys Rev B 90(4):045104. https://link.aps.org/doi/10.1103/PhysRevB.90.045104

# Diagrammatic Monte Carlo and GW Approximation for Jellium and Hydrogen Chain

# **20**

Kris Van Houcke, Igor S. Tupitsyn, and Nikolay V. Prokof'ev

## Contents

## Abstract

Within the general framework of skeleton diagrammatic expansions, the fully self-consistent GW approximation (sc-GW) is the simplest scheme based on the lowest-order diagrams. However, this established method for electronic structure calculations is rarely used in its original form when the polarization function is obtained from the product of two fully dressed single-particle Green's functions because its most serious deficiency is known to be an incorrect prediction of the dielectric response. In this contribution, we examine the sc-GW approximation

K. Van Houcke
Laboratoire de Physique Statistique, Ecole Normale Supérieure, UPMC, Université Paris Diderot, CNRS, Paris, France
e-mail: kris.van.houcke@lps.ens.fr

I. S. Tupitsyn (✉) · N. V. Prokof'ev
Department of Physics, University of Massachusetts, Amherst, MA, USA
e-mail: itupitsyn@physics.umass.edu; prokofev@physics.umass.edu

for the homogeneous electron gas and find that problems with the dielectric response are solved by enforcing the particle number conservation law in the polarization function. Our protocol for restoring physical properties of the sc-GW approximation (and more advanced schemes based on computing higher-order vertex corrections) is physically transparent and easy to implement at no additional computational cost. We further examine the accuracy of the sc-GW approximation and systematic convergence of the bold diagrammatic Monte Carlo scheme to the exact result for the single-orbital hydrogen chain system.

# 1    Introduction and Formalism

Dealing with interacting many-electron systems and accurately predicting properties of real solid-state materials are the major challenges of great technological importance. Among numerous theoretical methods developed for this problem, the diagrammatic many-body perturbation theory (Fetter 1971; Mahan 2000) offers a number of unique advantages. Instead of dealing with finite ensembles of electrons, the field-theoretical approach discussed in this work is based on the standard Feynman diagrammatic technique formulated in terms of single-particle propagators, or Green's functions, and interactions. Even the lowest-order schemes can grasp the essential physics of screening effects and allow one to deal with the long-range Coulomb interaction in the thermodynamic limit. One of the simplest schemes is the random-phase approximation (RPA) that can qualitatively (and even quantitatively) explain such features of metals as screening, plasmon, and Friedel oscillations. In principle, the skeleton diagrammatic expansion discussed below allows one to systematically improve on these results and obtain accurate solutions to the many-electron problem. In practice, however, progress is hindered because more sophisticated than RPA lowest-order diagrammatic approximations are not necessarily leading to better results right away (and in this contribution, we explain why and how this is happening), while evaluation of higher-order vertex corrections within the skeleton approach appears to be too costly computationally (apart from important questions concerning series convergence).

For three-dimensional systems with Coulomb interactions, one cannot use the diagrammatic expansion in terms of the bare interaction potential $V(k) = 4\pi e^2/k^2$ because of divergencies coming from geometric series-based "bubble" diagrams. The way out is to work with a formulation that expresses all diagrams self-consistently in terms of fully dressed propagators and interactions (Hedin 1965). Here we employ the skeleton expansion of the Luttinger-Ward (LW) functional $\Phi[G, W]$ (Luttinger and Ward 1960). The method is thermodynamically consistent and conserving; in particular, the relation between the Fermi momentum $k_F$ and particle density is preserved (Baym and Kadanoff 1961; Baym 1962). As the corresponding diagrams for the electron self-energy, $\Sigma$, and polarization, $\Pi$, are composed of three-point vertices connected pairwise by fully dressed Green's functions, $G$, and screened effective interactions, $W$ (with the exception of the Hartree diagram that is based on $V$), the corresponding diagrammatic technique

is often called the $G^2W$-expansion. Self-consistency is achieved through the Dyson equations written symbolically as

$$G = G_0 + G_0 \Sigma G, \qquad\qquad W = V + V \Pi W, \qquad\qquad (1)$$

where $G_0$ is the bare electronic propagator (depending on the representation, the products on the r.h.s. should be understood as convolutions over the space, time, spin, orbital, etc. indexes not mentioned here for brevity).

The self-consistent GW (sc-GW) approximation truncates the skeleton sequence at the lowest-order graphs (see upper left panel in Fig. 1), and contributions to $\Sigma$ and $\Pi$ are based on pair products of the $G$ and $W$ functions. Within the bold diagrammatic Monte Carlo (BDMC) framework (see, e.g., Van Houcke et al. 2012; Kulagin et al. 2013), the configuration space of leading vertex corrections (upper right panel in Fig. 1) and higher-order diagrams (lower panel in Fig. 1) is sampled stochastically. Since different diagram topologies contribute with a different sign, one might naively expect to be faced with a sign problem in the Monte Carlo simulation. Note, however, that there is no conventional sign problem:

(i) the diagrammatic theory is formulated directly in the thermodynamic limit, and therefore there is no exponential scaling of the computational time with system volume (Prokof'ev and Svistunov 2007; Van Houcke et al. 2012);

(ii) the sign alternation of diagrammatic contributions is essential to be able to sum the full series: in the best-case scenario, it can render the series convergent, and polynomial computational complexity can be achieved (Rossi 2017; Rossi et al. 2017) – in this specific sense, it is rather "sign blessing" than problem;

(iii) there is no vanishing denominator to deal with because normalization is established through the lowest-order terms.
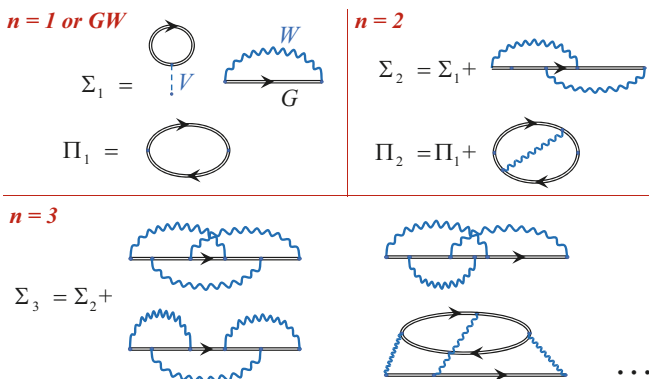


**Fig. 1** Skeleton diagrams for self-energy and polarization function in terms of fully dressed Green's functions (double lines) and screened interactions (wavy lines) at order $n = 1$ (equivalent to the sc-GW approximation) and order $n = 2$, upper left and right panels, respectively. In the lower panel, we show some of the self-energy diagrams at order $n = 3$

Moreover, diagrammatic Monte Carlo can be applied to any system at finite temperature with arbitrary dispersion relation (doped or undoped) and with arbitrary shape of the interaction potential. Both $\Sigma$ and $\Pi$ are computed as sums of skeleton graphs, up to order $n$; we denote these sums as $\Sigma_n$ and $\Pi_n$ and abbreviate the corresponding level of approximation as $BDMC_n$, i.e., $BDMC_1$ is identical to sc-GW. To obtain final answers, one has to perform an extrapolation to the $n \to \infty$ limit; this was demonstrated for several Coulomb systems in Tupitsyn et al. (2016) and Tupitsyn and Prokof'ev (2017).

In this contribution, we discuss two related topics. First we examine properties of the sc-GW approximation for the jellium model – a system of electrons on a homogeneous positive charge background with the same charge density – and present a simple strategy to enforce physical behavior for charge correlations. Our strategy can be applied at every order of the skeleton expansion and does not produce any systematic bias in the infinite-order limit for convergent series. We also provide accurate values for the ground-state energy, the quasiparticle $Z$-factor, and the effective mass renormalization $m_*/m$ (where $m$ is the bare electron mass) at the Fermi level. Second, we discuss merits of the sc-GW approximation for the single-orbital hydrogen chain system and demonstrate that the BDMC method allows one to obtain highly accurate results with full control over systematic errors by evaluating higher-order corrections. Most of the discussion closely follows material presented in Van Houcke et al. (2017) and Motta et al. (2017).

## 2    Jellium Model

Self-consistent GW approximation is the most established "beyond RPA" diagrammatic method for electronic structure calculations (Hedin 1965; Aryasetiawan and Gunnarsson 1998; Onida et al. 2002). However, by evaluating "bubble" diagrams in a self-consistent way [we do not discuss here numerous variations of the method based on partial, or incompletely dressed, propagators and interaction lines], the sc-GW approximation fails to reproduce some key results for the two-particle correlation functions and does not properly describe even the plasmon properties, in contrast to RPA. This drawback has been clearly demonstrated for a homogeneous electron gas (jellium model) by Holm and von Barth in Holm and von Barth (1998). Moreover, incorrect screening properties are expected to have a feedback on single-particle spectra of real materials for which the sc-GW approximation sometimes fails to account for the observed value of the absolute bandgap (Schöne and Eguiluz 1998).

Let us start by briefly reviewing the sc-GW approximation for the unpolarized jellium model. In the position-imaginary time $(r, \tau)$-representation, the lowest-order skeleton diagrams for the self-energy $\Sigma$ and polarization $\Pi$ read as (the Hartree diagram is canceled by the positive background)

$$\Sigma(r, \tau) = -G(r, \tau)W(r, -\tau), \qquad \Pi(r, \tau) = 2G(r, \tau)G(r, -\tau), \qquad (2)$$

where the factor of two comes from the sum over the spin index. These are self-consistently defined through algebraic solutions of the Dyson equations in the momentum-Matsubara frequency $(k, \omega_n)$ representation:

$$G(k, \omega_n)^{-1} = G^0(k, \omega_n)^{-1} - \Sigma(k, \omega_n), \qquad W(k, \omega_n)^{-1} = V(k)^{-1} - \Pi(k, \omega_n).$$
(3)

Here $G^0_\sigma = (i\omega_n + \mu - k^2/2m)^{-1}$. Knowing the one-body Green's function $G$ is sufficient for obtaining the system's energy, as well as quasiparticle properties such as $m_*$ and $Z$ (see Fetter 1971; Mahan 2000).

We performed all calculations at finite temperatures well below the Fermi energy $\varepsilon_F = k_F^2/2m$. Equations (2) and (3) were solved by iterations until the convergence was reached. The chemical potential $\mu$ used in each step was adjusted to preserve the desired electron density with accuracy up to six significant digits. After convergence, the total energy is computed from $\Sigma$ and $G$. For ground-state properties, we extrapolated results to zero temperature using the Fermi-liquid behavior. In Fig. 2, we show a typical plot for energy measured in units of Hartree, $Ha = me^4/\hbar^2$ at $r_s = 1$, with standard definition of $r_s$ as the ratio of the typical interparticle spacing and the Bohr radius $a_B = \hbar^2/me^2$.

## 2.1    Dielectric Response

The work by Holm and von Barth (1998) has established that the sc-GW approximation is not suitable for analysis of two-particle correlation functions. More precisely, the spectral function $S(k, \omega)$ of the polarization $\Pi$ was found to have incorrect
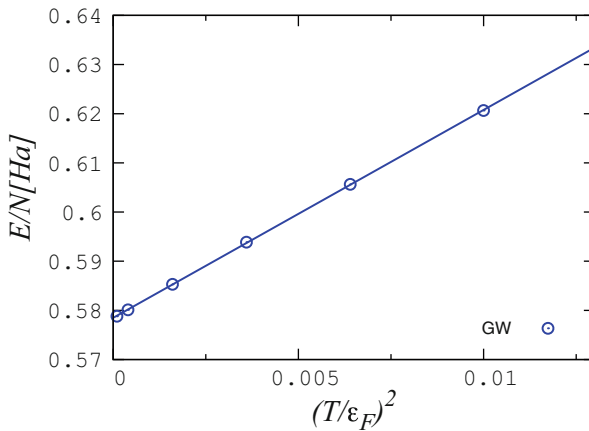


**Fig. 2** Energy per electron (in Ha) as a function of $(T/\varepsilon_F)^2$ revealing the Fermi-liquid behavior. The solid line is a linear fit giving a ground-state energy of $E/N = 0.5783(2)$ Ha
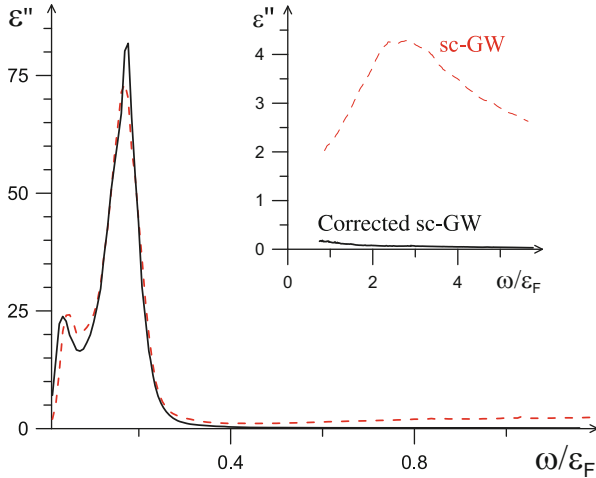
**Fig. 3** Imaginary part of the dielectric function within the sc-GW approximation at $r_s = 1$, $k/k_F = 0.1$, and $T/\varepsilon_F = 0.02$. Red dashed curve is the original sc-GW result, and the solid black line is the corrected sc-GW spectrum. The crucial difference at frequencies $\omega > k v_F$ is clearly seen in the inset

behavior at frequencies $\omega > k v_F$, where $v_F$ is the Fermi velocity; as a consequence, the real part of the dielectric function $\varepsilon(k, \omega) = 1 - (4\pi e^2/k^2)\Pi(k, \omega)$ at small momenta $k \ll k_F$ has its zero shifted away from the plasmon frequency $\omega_p = \sqrt{4\pi n e^2/m}$ to completely unrealistic values (see Fig. 3 in Holm and von Barth (1998)).

Our results confirm this key observation: we also find that at $k \ll k_F$ and $\omega_n \gg k v_F$, the polarization is *orders of magnitude* larger than the values dictated by the hydrodynamic plasmon mode physics, $\Pi(k, \omega_n) \approx -nk^2/m\omega_n^2$ (see thorough discussion in Nozieres (1999)). This unphysical behavior can be traced back to the fact that the sc-GW approximation does not respect the dynamic particle number conservation law, which implies that at zero momentum, $\Pi(k = 0, \omega_n) \propto \delta_{n,0}$ or, identically, $\Pi(k = 0, \tau) = \text{const}$. Indeed, for an arbitrary interaction potential, $\Pi$ is related to the density-density correlation function, $\chi(k, \omega_n) = \langle |\delta \hat{n}(k, \omega_n)|^2 \rangle$, as $\Pi = -\chi/(1 - V\chi)$, while (in imaginary-time representation) $\chi(k = 0, \tau) = \langle \delta \hat{N}(0) \delta \hat{N}(\tau) \rangle \equiv \text{const}$ because the total number of particles $\hat{N}$ commutes with the Hamiltonian. [To prevent divergence of $V(k)$ when substituting $k = 0$, the bare potential may be regularized at small momenta to have the Yukawa form, $V(k) = \frac{4\pi e^2}{k^2 + k_0^2}$, with small but finite Yukawa screening wave vector $k_0$ that is sent to zero at the end of the calculation.] As a result, at $k = 0$ the density response is purely static, $\chi(k = 0, \omega_n \neq 0) \equiv 0$ (see Nozieres 1999), implying that the same is true for the polarization function

$$\Pi(k = 0, \omega_n \neq 0) \equiv 0, \qquad \text{(Particle number conservation law).} \qquad (4)$$

By assuming physical behavior of $\Pi$ at small momenta, one would conclude that $W/V - 1 \to -\omega_p^2/(\omega_p^2 + \omega_n^2)$ and frequencies larger than $100\omega_p$ are not required for highly accurate calculations.

Instead, within the sc-GW approximation, one finds that $\Pi(k = 0, \omega_n)$ has significant amplitudes at finite frequencies, and correspondingly, $\Pi(k, \omega_n \neq 0)$ is not approaching zero when $k \to 0$. This also causes significant problems for the proper technical implementation of the sc-GW approach in Coulomb systems because $(4\pi e^2/k^2)\Pi(k, \omega_n \neq 0)$ tends to diverge as small momenta and forces one to consider extremely large frequencies (exceeding $10^6 \omega_p$) in the calculation of the screened interaction $W$. We believe that this frequency range was not covered in the sc-GW implementations reported in Holm and von Barth (1998), Holm (1999), and García-González and Godby (2001) leading to the discrepancy in the final results outside of reported error bars (see Table 1).

Since all problems originate from the violation of the dynamic particle conservation law, we propose a simple strategy to enforce the physical behavior of $\Pi(k, \omega_n)$. All one has to do is to perform a transformation:

$$\Pi(k, \omega_n \neq 0) \to \Pi(k, \omega_n) - \Pi(0, \omega_n \neq 0), \qquad (5)$$

before calculating the dielectric response from the sc-GW solution. In other words, one has to subtract the spurious frequency dependence at $k = 0$. Note that this transformation is compatible with the higher-order skeleton expansion and the suggestion is that it should be implemented *within* the fully self-consistent skeleton scheme whenever one has to iterate properties of the $W$-function. Indeed, in the large-order expansion limit, the correction term is supposed to vanish as $\Pi(k = 0, \omega_n)$ converges to the correct physical behavior $\propto \delta_{n,0}$.

In Figs. 3 and 4, we show how our protocol works by considering the case of $r_s = 1$ at low temperature $T/\varepsilon_F = 0.02$ and small momentum $k/k_F = 0.1$. After performing numeric analytic continuation of the imaginary frequency data for $\varepsilon(k, \omega_n)$ to get the imaginary part $\varepsilon''(k, \omega)$ (using protocols described in Goulko et al. 2017; Mishchenko 2012), we obtain the real part $\varepsilon'(k, \omega)$ from the Kramers-Kronig relation. The improvement in terms of eliminating the unphysical behavior is dramatic. After the transformation, the high-frequency tail of $\varepsilon''(k, \omega)$ gets suppressed by nearly two orders of magnitude. As a result, the real part of the dielectric function now has its zero at $\omega_p^{(GW)} \approx 0.89(1)\omega_p$ and is approaching unity from below at $\omega \gg \varepsilon_F$. [In order to have $\omega_p^{(GW)}$ to coincide with $\omega_p$ within the sc-GW approximation, one would need to divide $\Pi$ by $Z^2$ by hand, mimicking the effect of vertex corrections that can be obtained properly only by implementing the BDMC$_n$ scheme.] Everything about the original sc-GW data at frequencies $\omega > k v_F$ is completely unsatisfactory.

**Table 1** Minus the ground-state exchange-correlation energy per particle $-E_{XC}$ (in Hartree). We compare results for $-E_{XC}$ obtained within the sc-GW approximation discussed in this chapter with quantum Monte Carlo data from Ceperley and Alder (1980) (labeled as QMC1) and Ortiz et al. (1999) and Ortiz and Ballone (1994) (labeled as QMC2) and with the earlier implementations of the sc-GW approximation data published in García-González and Godby (2001) (labeled as GW1) and Holm and von Barth (1998) (labeled as GW2)

| $r_s$ | 1 | 2 | 4 | 5 | 10 |
|---|---|---|---|---|---|
| QMC1 | 0.5180 | 0.2742 | 0.1464 | 0.1197 | 0.0644 |
| QMC2 | 0.5127 | 0.2713 | | 0.1201 | |
| GW2 | | 0.2741 | 0.1465 | | |
| GW1 | 0.5160(2) | 0.2727(5) | 0.1450(5) | 0.1185(5) | 0.0620(9) |
| GW (this work) | 0.5267(2) | 0.2789(1) | 0.1488(1) | 0.1216(1) | 0.06498(2) |
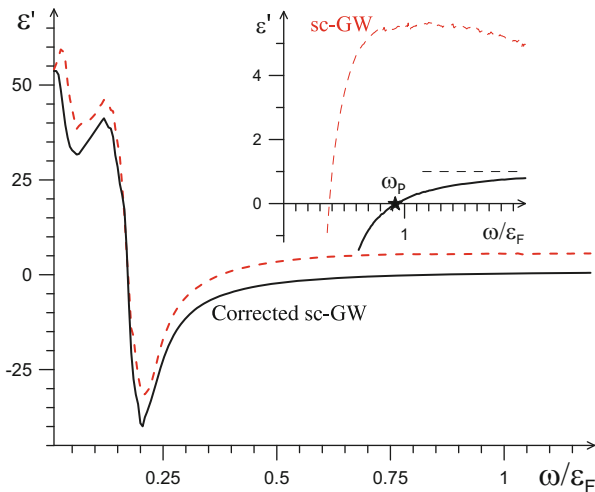


**Fig. 4** Real part of the dielectric function within the sc-GW approximation at $r_s = 1$, $k/k_F = 0.1$, and $T/\varepsilon_F = 0.02$. The original sc-GW result (red dashed curve) completely misses the plasmon zero and predicts unrealistically large response at frequencies above $\varepsilon_F$. The corrected result (solid black line) crosses zero within 10% of $\omega_p$ and saturates to unity at $\omega > \varepsilon_F$

## 2.2    Ground-State Properties

Since precise knowledge of $G_\sigma$ is required in order to calculate $\Pi$ via Eq. (2), we extract basic Fermi-liquid properties from the Green's function and compare them against known results. Early implementations of the sc-GW approximation (García-González and Godby 2001; Holm and von Barth 1998; Holm 1999) reported ground-state energy values in very good agreement (often at the sub-percent level, see Table 1) with diffusion Monte Carlo results (Ceperley and Alder 1980). We find that our sc-GW exchange-correlation energies differ from those of García-González and Godby (2001), Holm and von Barth (1998), and Holm (1999) well

**Table 2** The quasiparticle residue $Z$ and the effective mass renormalization $m_*/m$ at the Fermi level for the unpolarized 3D homogeneous electron gas. These quantities were obtained by using the *standard* sc-GW approach, i.e., solving the set of Eqs. (2) and (3) without any correction procedures to cure the unphysical behavior of $\Pi$

| $r_s$ | 1 | 2 | 4 | 5 | 10 |
|---|---|---|---|---|---|
| $Z$ | 0.899(1) | 0.842(1) | 0.769(2) | 0.743(2) | 0.658(2) |
| $m_*/m$ | 0.944(2) | 0.931(2) | 0.913(2) | 0.906(2) | 0.875(2) |

**Table 3** Minus the ground-state exchange-correlation energy per particle $-E_{XC}$ (in Hartree), the quasiparticle residue $Z$, and the effective mass renormalization $m_*/m$ at the Fermi level for the unpolarized 3D homogeneous electron gas. These are obtained by using the *modified* sc-GW, or GW($\Pi$), approach, i.e., by solving the set of Eqs. (2) and (3) self-consistently and at the same time applying the transformation Eq. (5) to $\Pi$ at each iteration of the self-consistent scheme.

| $r_s$ | 1 | 2 | 4 | 5 | 10 |
|---|---|---|---|---|---|
| $-E_{XC}$ | 0.5205(2) | 0.2716(2) | 0.1413(1) | 0.1143(1) | 0.05873(2) |
| $Z$ | 0.880(3) | 0.808(2) | 0.722(2) | 0.692(3) | 0.605(4) |
| $m_*/m$ | 0.934(3) | 0.900(2) | 0.834(2) | 0.803(2) | 0.673(4) |

outside the error bounds by an amount bigger than the difference between the sc-GW and various other approximations, for instance GW$^{(0)}$. While the exact origin of discrepancy is not settled yet, it is most likely linked to the violations of the particle conservation law (4) discussed above. To ensure correctness of our sc-GW results, we developed two absolutely independent codes that did not share a single common idea about grids and cutoffs for storing and processing the data, Fourier transforms, and energy evaluation (see next subsection). It turned out that energy data used for plots in Yan (2011) were also in perfect agreement with our results; ultimately, five independently developed finite-$T$ codes (including codes developed by L. Pollet and S. Schultess) were compared and found to be in agreement with each other within the error bounds reported in Table 2.

For benchmark purposes, we also report the quasiparticle residue at the Fermi surface, or $Z$-factor, and the effective mass renormalization in Table 2. Error bounds were estimated from variations induced by changing momentum-time grids, cutoffs, and extrapolation procedures to the zero-temperature limit. All results in the table were obtained for the *standard* sc-GW formulation; i.e., the transformation procedure (5) was *not* applied when solving Eqs. (2) and (3).

While applying the transformation Eq. (5) to the sc-GW solution vastly improves the two-body spectral properties, it is natural to ask what impact it has on the Fermi-liquid properties when applied at each iteration of the self-consistent scheme; we abbreviate the corresponding scheme that always respects (4) as GW($\Pi$). The exchange-correlation energy, $Z$-factor, and effective mass renormalization obtained in that way are given in Table 3. The relative change in exchange-correlation energy ranges from about 1% for $r_s = 1$ up to almost 10% for $r_s = 10$. Given that the relative difference in exchange-correlation energy calculated within the standard sc-GW approach and the diffusion Monte Carlo method (Ceperley and Alder 1980)
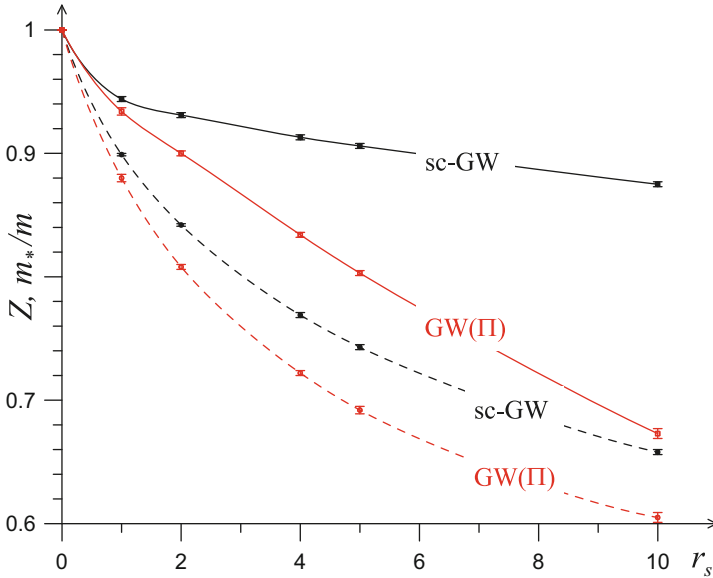
**Fig. 5** Effective mass renormalization (solid curves) and quasiparticle residue (dashed curves) as functions of $r_s$ within the sc-GW approach (black) and its modified GW($\Pi$) version

is at the level of few percent, we conclude that the GW($\Pi$) scheme has the same quality in terms of the ground-state energy while it improves the two-body spectral function by orders of magnitude.

The effect of the transformation (5) on $Z$ and especially on $m_*/m$ is more pronounced (see Fig. 5). If changes for the quasiparticle residue can be considered "quantitative," the same cannot be said about the effective mass renormalization – it is far more dramatic within the GW($\Pi$) approach. Moreover, the dependence of $m_*/m$ on $r_s$ within the sc-GW approximation can hardly be extrapolated to unity for the noninteracting system without assuming an anomalously strong increase in the $dm_*/dr_s$ derivative.

## 2.3    Details on Practical Implementation

To verify the correctness of treating the properties at small momenta in our implementation of the sc-GW scheme (no matter how unphysical they might be), a good strategy is to develop a code for the screened Yukawa potential as well. Let us refer to these implementations as the Coulomb code and the Yukawa code, respectively. To recover the final answer for the Coulomb system with the Yukawa code, one should extrapolate results obtained for a set of small $k_0$ values to zero (see Fig. 6). The expectation is that the function is linear. Let us highlight some technical details of the two implementations. We discuss the Yukawa code first.

**Fig. 6** Energy per electron (in Hartree) as a function of Yukawa screening wavevector for $T/\varepsilon_F = 0.02$ and $r_s = 1$. The result of the Coulomb code is also shown
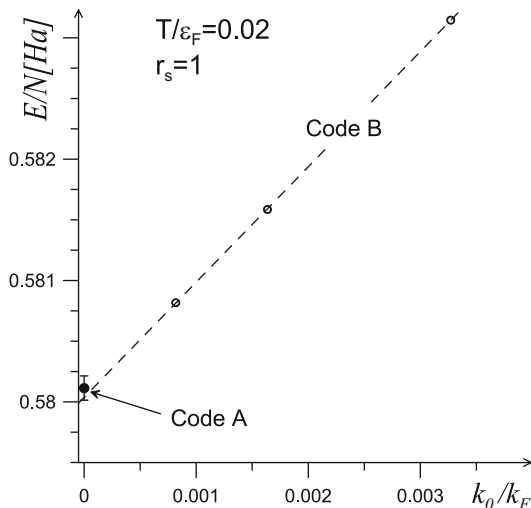


In the Yukawa code, the high-momentum cutoff was set at $17k_F$ for the Green's function and at $33k_F$ for the screened potential. We verified that by increasing these cutoffs by 50%, energies and other Fermi-liquid parameters remained the same up to six significant digits. The smallest momentum resolution was set at $10^{-4}k_F$. This resolution was used in the vicinity of three special points: $k = 0$, $k_F$, and $2k_F$. Away from the special points, the grids were made more sparse (we used logarithmic grids with 16 equidistant points per scale). Similarly, in imaginary-time domain, the smallest time resolution was set at $5 \times 10^{-5}/\varepsilon_F$ in the vicinity of two special points $\tau = 0$ and $\tau = \beta$, and we used the same law (16 equidistant points per scale) to describe the entire domain. All functions in the momentum-imaginary time representation were obtained by considering parabolic (in both momentum and time) interpolations between the grid points. When solving the Dyson equation, we considered up to $2^{20}$ Matsubara frequencies. Our parameters were such that by decreasing the resolution of the grids (increasing the number of grid points) by a factor of two, or increasing the number of Matsubara frequencies, we were not able to detect changes in the final results within the reported error bounds.

When computing the self-energy and polarization function values (in imaginary-time representation) from the momentum integrals of the $GW$ and $GG$ products, the Yukawa code was relying on the direct numerical integration, i.e., the Fourier transforms to the real-space and back to momentum space were not used.

In the Coulomb code, on the other hand, the high-momentum cutoff was set at $25k_F$, for both the Green's function and the screened potential, and it was also verified that final results did not change upon increasing this cutoff. A logarithmic grid was used for small momenta (the smallest momentum being $10^{-8}k_F$). In the region $[0.1k_F, 2k_F]$, the typical spacing is of the order $10^{-3}k_F$, while for momenta bigger than $2k_F$, it is of the order $10^{-2}k_F$. In imaginary-time domain, the smallest time resolution was set at $10^{-6}/\varepsilon_F$ in the vicinity of $\tau = 0$ and $\tau = \beta$, and we used

the same law (80 equidistant points per scale) to describe the entire domain. For the frequency domain, a logarithmic grid was used, the largest frequency being of the order $10^7 \times \pi/\beta$. No dependence of the final results on this cutoff was observed.

In contrast to the Yukawa code, the convolutions to obtain the self-energy and polarization function values in imaginary-time representation were done by explicitly going to position space and back to momentum space via (slow) numerical Fourier transforms. The grid in position space was logarithmic with cutoff $500/k_F$, and the smallest position grid point corresponded to $10^{-4}/k_F$.

## 3  Hydrogen Chain

As it was already mentioned, the sc-GW approximation is one of the most established methods for electronic structure calculations. It is widely used in material science systems, but in many cases, its systematic bias remains unknown. Below we address this question within the hydrogen chain problem (Hachmann et al. 2006) using the BDMC technique. Despite its simplicity, the hydrogen chain contains all the basic ingredients distinguishing real materials from the jellium model.

### 3.1  Hamiltonian and Diagrammatic Representation

In atomic units, the chain of hydrogen atoms, composed of N fixed equidistant protons and N electrons, is described by the Hamiltonian:

$$H = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{i,m} \frac{1}{|\mathbf{r}_i - \mathbf{R}_m|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2} \sum_{m \neq n} \frac{1}{|\mathbf{R}_m - \mathbf{R}_n|}, \quad (6)$$

where the first two terms give the kinetic and potential energy of $i$-th electron in the protons' field while the third and fourth terms represent the electron-electron and proton-proton Coulomb interactions. The positions of electrons and protons are described by $\mathbf{r}_i$ and $\mathbf{R}_m$, respectively. The protons along the chain are separated by the distance $R$.

We start by projecting this Hamiltonian onto the finite orbital space using field operators:

$$\Psi_\sigma(\mathbf{r}) = \sum_{i,\alpha} \phi_{i,\alpha,\sigma}(\mathbf{r}) \, f_{i,\alpha,\sigma}, \quad (7)$$

where $\phi_{i,\alpha,\sigma}(\mathbf{r}) \equiv \phi_{\alpha,\sigma}(\mathbf{r} - \mathbf{R}_i)$ is the $\alpha$-th orbital of the electron with spin $\sigma$ at the $i - th$ proton and $f_{i,\alpha,\sigma}$ is the standard electron annihilation operator. To simplify notations, we will also use a composite index $a = (i, \alpha, \sigma)$, as in Fig. 7. In terms of these operators, the Hamiltonian can be presented as
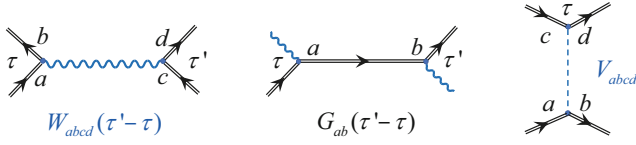
**Fig. 7** Diagrammatic elements representing screened interaction, $W$, Green's function, $G$, and bare interaction, $V$, in the orbital representation. Each index $\{a, b, c, d\}$ combines all three discrete indices $(i, \alpha, \sigma)$

$$\hat{H} = -\sum_{a,b} t_{a,b}\delta_{\sigma_b,\sigma_a}\, f_a^\dagger\, f_b + \frac{1}{2}\sum_{a,b,c,d} V_{a,b,c,d}\delta_{\sigma_a,\sigma_b}\delta_{\sigma_c,\sigma_d}\, f_a^\dagger\, f_c^\dagger\, f_d\, f_b \qquad (8)$$

with nonzero hopping amplitudes and interactions given by

$$t_{a,b} = -\int d\mathbf{r}\, \phi_{i,\alpha,\sigma}^*(\mathbf{r})\left[-\frac{1}{2}\nabla^2 - \sum_a \frac{1}{|\mathbf{r} - \mathbf{R}_a|}\right]\phi_{j,\beta,\sigma}(\mathbf{r}) \qquad (9)$$

and

$$V_{a,b,c,d} = \int d\mathbf{r}d\mathbf{r}'\, \phi_{i,\alpha,\sigma}^*(\mathbf{r})\, \phi_{k,\gamma,\sigma'}^*(\mathbf{r}')\, \frac{1}{|\mathbf{r} - \mathbf{r}'|}\, \phi_{l,\delta,\sigma'}(\mathbf{r}')\, \phi_{j,\beta,\sigma}(\mathbf{r}). \qquad (10)$$

All integrals are taken over the three-dimensional orbital coordinates. In this representation, electronic coordinates are replaced with discrete orbital indices. This solves the problem of strong coupling between the electrons and ions and leads to the diagrammatic technique that is more or less standard for lattice models. Of course, part of the problem complexity is now in the size of the orbital Hilbert space.

From the set of Eqs. 8, 9, and 10, it is clear that the topology of diagrams within the $G^2W$ skeleton expansion remains unchanged, but their index structure gets more complicated. To be more specific, in Fig. 7, we show the structure of all diagrammatic elements in the orbital representation. Each interaction line depends on four site/atom indices $\{(i, j); (k, l)\}$, four orbital indices $\{(\alpha, \beta); (\gamma, \delta)\}$, and two spin indices $\{\sigma; \sigma'\}$ (note that the Coulomb interaction vertex does not change the spin index).

In what follows, we further simplify the problem and limit ourselves to considering one orbital per atom (the so-called minimal basis set; STO-6G basis). In this case, the Hamiltonian takes the form

$$\hat{H} = -\sum_{i,j}\sum_{\sigma} t_{i,j}(\sigma)\, f_{i,\sigma}^\dagger\, f_{j,\sigma} + \frac{1}{2}\sum_{i,j,k,l}\sum_{\sigma,\sigma'} V_{i,j,k,l}(\sigma, \sigma')\, f_{i,\sigma}^\dagger\, f_{k,\sigma'}^\dagger\, f_{l,\sigma'}\, f_{j,\sigma}$$

$$(11)$$

and the 1st order contributions to the self-energy $\Sigma$ and polarization $\Pi$ become (we employ periodic boundary conditions)

$$\Sigma_H(|j-i|, \sigma) = \sum_{\sigma'} \sum_{k,l} G(|k-l|, \tau = -0, \sigma') \, V_{i,j,k,l}(\sigma, \sigma') \quad (12)$$

$$\Sigma_{GW}(|l-i|, \tau, \sigma) = -\sum_{j,k} G(|k-j|, \tau, \sigma) \, W_{i,j,k,l}(-\tau, \sigma, \sigma) \quad (13)$$

$$\Pi(i, j, k, l, \tau, \sigma, \sigma) = G(|k-j|, \tau, \sigma) \, G(|i-l|, -\tau, \sigma), \quad (14)$$

where $G$ and $W$ are the solutions of the Dyson equations. To avoid double counting, the Hartree diagram with fully dressed Green's function should contain only the bare (time-independent) interaction $V_{i,j,k,l}(\sigma, \sigma')$.

To make use of the translation invariance, we introduce variables $u = i - j$ and $v = k - l$, defining relative distances between the orbital centers. Then $u = v = 0$ corresponds to the "density-density" part of the interaction potential $V_{i,j,k,l}(\sigma, \sigma')$. For the model under consideration, this density-density part dominates the answer, and inclusion of terms with nonzero values of $u$ and $v$ in the Dyson equation for $W$ introduces only minor changes. We find that energies per atom, computed with cutoffs $u^* = v^* = 2$, differ from the full answer just by $\sim 10^{-5}$ in relative units even at the smallest inter-atom separations. The possibility to introduce such cutoffs in a controlled way enormously simplifies the computational effort and memory requirements in high diagrammatic orders.

## 3.2 Pauli Exclusion Principle: Zero Terms

There exist two different standards in the formulation of the projected Hamiltonian. The first approach explicitly respects the Pauli exclusion principle in the interaction terms, while the second approach pays no special attention to it. Physically, the two standards deal with identical Hamiltonians and would lead to the same answer if solved exactly. However, if one truncates self-consistent diagrammatic series at low order (as in the sc-GW approximation), the two standards of formulating the projected Hamiltonian would produce different answers. Given that these two approaches are "inherent" to two different communities (lattice models *vs* material science) and that the sc-GW approximation is a widely used method, we consider both alternative here to quantify the effect of adding/subtracting zero Hamiltonian terms on the sc-GW approximation.

To illustrate the point, consider an ideal spin-polarized lattice Fermi gas with added contact interaction term

$$\hat{H}_0 = -\sum_{ij} t_{ij} \, f_{j\uparrow}^\dagger \, f_{i\uparrow} + U \sum_i f_{i\uparrow}^\dagger \, f_{i\uparrow}^\dagger \, f_{i\uparrow} \, f_{i\uparrow}. \quad (15)$$

**Fig. 8** Effect of zero terms on the sc-GW results for the hydrogen chain: energy per atom (minimal basis set, thermodynamic limit). Black line: the DMRG curve (best variational answer). Blue line: the sc-GW($H$) curve (the material science protocol). Red line: the sc-GW($H'$) curve (the lattice model protocol)
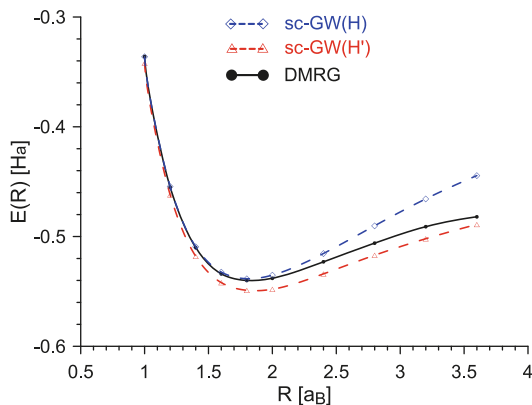


Because of the Pauli principle, this interaction term is zero. Even if we keep it, in the diagrammatic expansion with respect to $U$, all diagrams of the same order will cancel each other, order by order. However, this is true only in the case of exact solution accounting for all the orders in $U$. The sc-GW scheme includes some (but not all) higher-order corrections in $U$; unaccounted terms will thus lead to different answers for Eq. 15 with $U = 0$ and $U \neq 0$.

Similar considerations hold true for any Hamiltonian projected onto the orbital basis. The terms with matrix elements $V_{abad}$ and $V_{bada}$ have no effect on the exact answer because of the Pauli principle – they are "zero terms" and one can either (i) keep them (material science community protocol) or (ii) drop them (lattice model Hamiltonian community protocol). We distinguish these two protocols by introducing notation (i) $\hat{H}$ and (ii) $\hat{H}'$, respectively. In Fig. 8 (energy per atom for different values of the lattice distance $R$), we demonstrate how the same sc-GW scheme for two formally identical Hamiltonians $\hat{H}$ and $\hat{H}'$ produces different answers. As one can see from this figure, the effect of zero terms on the sc-GW results is far from being negligible, especially in the region of large inter-atom distances $R > 1.8 \ a_B$. Also, while the sc-GW($\hat{H}$) is more accurate in the perturbative region ($R < 1.8 \ a_B$), the sc-GW($\hat{H}'$) appears to produce more consistent energies throughout the entire range of $R$. Since the two sc-GW answers surround the DMRG one, the difference between them can be considered as an estimate of the systematic bias within the sc-GW approximation. Indeed, in the next subsection, we show that this difference has similar magnitude as the energy shift produced by considering the leading vertex correction.

## 3.3    Energy per Atom: Convergence of BDMC Results

To go beyond the sc-GW approximation, we employ the BDMC technique and obtain the ground-state energy per atom (using the same $T \rightarrow 0$ extrapolation

**Fig. 9** Ground-state energy in different skeleton orders (minimal basis set, thermodynamic limit). By accounting for higher-order vertex corrections, the BDMC results converge to the best variational answer – DMRG (black curve). Red line: The $BDMC_1(H') = sc\text{-}GW(H')$ curve. Green line: The $BDMC_2(H')$ curve. Blue line: the converged $BDMC(H')$ result (see text)
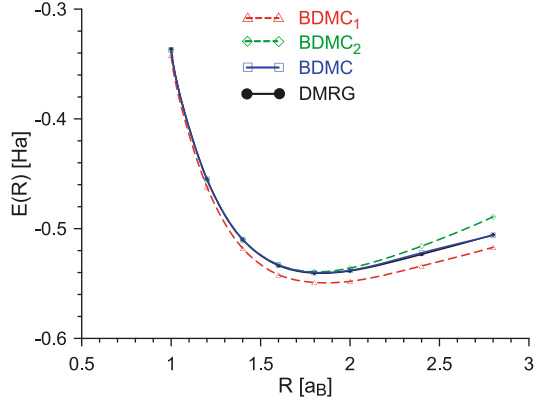
procedure as presented in Fig. 2) for system (11) in the thermodynamic limit (we work with $\hat{H}'$).

The convergence of the $BDMC_n$ results to the exact (DMRG) answer as the largest accounted for order $n$ is increased and is presented in Fig. 9. $BDMC_2$ data are already extremely accurate at the inter-atom separation $R \leq 1.8$ $a_B$. One has to increase the expansion order and consider $BDMC_3$ to achieve the same accuracy for $R = 2.4$ $a_B$, indicating that we are entering a non-perturbative regime. For $R = 2.8$ $a_B$, convergence is reached only when all diagrams up to the 5th skeleton order are included, while larger values of $R$ require simulation of even higher expansion orders and the development of appropriate resummation methods. From this study, we conclude that the BDMC technique for systems with Coulomb interactions can be used to obtain converged answers for interesting parameter regimes (see also Tupitsyn et al. 2016 and Tupitsyn and Prokof'ev 2017) with full control over systematic and statistical errors. The study of the hydrogen chain is the first application of the many-body diagrammatic technique to a material science system (Motta et al. 2017); usually such calculations are limited to the lowest-order schemes and involve an unknown systematic bias.

# 4    Conclusions

We have proposed a simple strategy to drastically improve key properties of the two-particle correlation functions within the sc-GW approximation and applied it to the jellium model. The strategy is designed to cure unphysical behavior of the polarization function $\Pi$ that originates from the violation of the dynamic particle conservation law. The very same trick can be applied to other models and materials science systems and can be used in the diagrammatic Monte Carlo approach that considers higher-order vertex corrections (which should correct the unphysical behavior of $\Pi$). We also provide benchmark values of key Fermi-liquid parameters

for jellium within the *standard* sc-GW approximation and the *modified* GW($\Pi$) version.

To explore merits of the bold diagrammatic Monte Carlo approach for materials, we applied this technique to a prototype material science system by evaluating diagrams of all relevant orders to obtain the answer with controlled accuracy. We proposed a protocol for estimating the systematic bias of the sc-GW method using an important ambiguity in the formulation of this diagrammatic approximation. It is related to two opposite ways of treating the Pauli exclusion principle and inherent to two different communities – to the lattice model Hamiltonian and material science ones. Being "innocent" at the level of exact solution, in the case of low-order schemes, such as the sc-GW approximation, this ambiguity leads to answers that differ by an amount similar to that coming from vertex corrections. We demonstrated that the BDMC technique for the single-orbital hydrogen chain converges to the exact answer, i.e., higher-order vertex corrections allow one to reach the desired accuracy in a controlled way.

# References

Aryasetiawan F, Gunnarsson O (1998) The *GW* method. Rep Prog Phys 61(3):237. http://stacks.iop.org/0034-4885/61/i=3/a=002

Baym G (1962) Self-consistent approximations in many-body systems. Phys Rev 127:1391–1401. https://link.aps.org/doi/10.1103/PhysRev.127.1391

Baym G, Kadanoff LP (1961) Conservation laws and correlation functions. Phys Rev 124:287–299. https://link.aps.org/doi/10.1103/PhysRev.124.287

Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45:566–569. https://link.aps.org/doi/10.1103/PhysRevLett.45.566

Fetter AL, Walecka JD (1971) Quantum theory of many-particle systems. McGraw-Hill, New York

García-González P, Godby RW (2001) Self-consistent calculation of total energies of the electron gas using many-body perturbation theory. Phys Rev B 63:075112. https://link.aps.org/doi/10.1103/PhysRevB.63.075112

Goulko O, Mishchenko AS, Pollet L, Prokof'ev N, Svistunov B (2017) Numerical analytic continuation: answers to well-posed questions. Phys Rev B 95:014102. https://link.aps.org/doi/10.1103/PhysRevB.95.014102

Hachmann J, Cardoen W, Chan GKL (2006) Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group. J Chem Phys 125:144101. http://aip.scitation.org/doi/full/10.1063/1.2345196

Hedin L (1965) New method for calculating the one-particle green's function with application to the electron-gas problem. Phys Rev 139:A796–A823. https://link.aps.org/doi/10.1103/PhysRev.139.A796

Holm B (1999) Total energies from *GW* calculations. Phys Rev Lett 83:788–791. https://link.aps.org/doi/10.1103/PhysRevLett.83.788

Holm B, von Barth U (1998) Fully self-consistent GW self-energy of the electron gas. Phys Rev B 57:2108–2117. https://link.aps.org/doi/10.1103/PhysRevB.57.2108

Kulagin SA, Prokof'ev N, Starykh OA, Svistunov B, Varney CN (2013) Bold diagrammatic monte carlo technique for frustrated spin systems. Phys Rev B 87:024407. https://link.aps.org/doi/10.1103/PhysRevB.87.024407

Luttinger JM, Ward JC (1960) Ground-state energy of a many-fermion system. II. Phys Rev 118:1417–1427. https://link.aps.org/doi/10.1103/PhysRev.118.1417

Mahan GD (2000) Many-particle physics. Springer, Boston

Mishchenko AS (2012) Correlated electrons: from models to materials. In: Pavarini E, Koch W, Anders F, Jarrel M (eds). Forschungszentrum Julich GmbH, Julich

Motta M, Ceperley DM, Chan GKL, Gomez JA, Gull E, Guo S, Jimenez-Hoyos C, Lan TN, Li J, Ma F, Millis AJ, Prokof'ev NV, Ray U, Scuseria GE, Sorella S, Stoudenmire EM, Sun Q, Tupitsyn IS, White SR, Zgid D, Zhang S (2017) Towards the solution of the many-electron problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods. Phys Rev X 7:031059. https://journals.aps.org/prx/abstract/10.1103/PhysRevX.7.031059

Nozieres P, Pines D (1999) Theory of quantum liquids, chapters 2 and 3. Westview Press, Boulder

Onida G, Reining L, Rubio A (2002) Electronic excitations: density-functional versus many-body Green's-function approaches. Rev Mod Phys 74:601–659. https://link.aps.org/doi/10.1103/RevModPhys.74.601

Ortiz G, Ballone P (1994) Correlation energy, structure factor, radial distribution function, and momentum distribution of the spin-polarized uniform electron gas. Phys Rev B 50:1391–1405. https://link.aps.org/doi/10.1103/PhysRevB.50.1391

Ortiz G, Harris M, Ballone P (1999) Zero temperature phases of the electron gas. Phys Rev Lett 82:5317–5320. https://link.aps.org/doi/10.1103/PhysRevLett.82.5317

Prokof'ev N, Svistunov B (2007) Bold diagrammatic Monte Carlo technique: when the sign problem is welcome. Phys Rev Lett 99:250201. https://link.aps.org/doi/10.1103/PhysRevLett.99.250201

Rossi R (2017) Determinant diagrammatic monte carlo algorithm in the thermodynamic limit. Phys Rev Lett 119:045701. https://link.aps.org/doi/10.1103/PhysRevLett.119.045701

Rossi R, Prokof'ev N, Svistunov B, Van Houcke K, Werner F (2017) Polynomial complexity despite the fermionic sign. EPL 118(1):10004. https://doi.org/10.1209/0295-5075/118/10004

Schöne WD, Eguiluz AG (1998) Self-consistent calculations of quasiparticle states in metals and semiconductors. Phys Rev Lett 81:1662–1665. https://link.aps.org/doi/10.1103/PhysRevLett.81.1662

Tupitsyn IS, Prokof'ev NV (2017) Stability of Dirac liquids with strong Coulomb interaction. Phys Rev Lett 118:026403. https://link.aps.org/doi/10.1103/PhysRevLett.118.026403

Tupitsyn IS, Mishchenko AS, Nagaosa N, Prokof'ev N (2016) Coulomb and electron-phonon interactions in metals. Phys Rev B 94:155145. https://link.aps.org/doi/10.1103/PhysRevB.94.155145

Van Houcke K, Werner F, Kozik E, Prokof'/ev N, Svistunov B, Ku MJH, Sommer AT, Cheuk LW, Schirotzek A, Zwierlein MW (2012) Feynman diagrams versus Fermi-gas Feynman emulator. Nat Phys 8(5):366–370. https://doi.org/10.1038/nphys2273

Van Houcke K, Tupitsyn IS, Mishchenko AS, Prokof'ev NV (2017) Dielectric function and thermodynamic properties of jellium in the $GW$ approximation. Phys Rev B 95:195131. https://link.aps.org/doi/10.1103/PhysRevB.95.195131

Yan XZ (2011) Approximation for discrete Fourier transform and application in study of three-dimensional interacting electron gas. Phys Rev E 84:016706. https://link.aps.org/doi/10.1103/PhysRevE.84.016706

Check for
updates

# Coupled Cluster and Quantum Chemistry Schemes for Solids

# 21

Andreas Grüneis

## Contents

**Abstract**

Quantum chemical wavefunction-based theories approximate the true many-electron wavefunction in a compact fashion and bear the potential to predict materials properties with high accuracy. Their computational complexity is significantly higher than that of the current workhorse method in computational materials science, density functional theory in the framework of approximate exchange, and correlation density functionals. However, the increase in available computer power and methodological and algorithmic improvements during the last decade have made quantum chemical studies of increasing system sizes possible. Coupled cluster theories are among the most widely used quantum chemical wavefunction-based methods. They employ an exponential ansatz of the electronic wavefunction that constitutes a good trade-off between accuracy and computational cost for weakly correlated many-electron systems. Here, we discuss methodological aspects and recent developments of coupled cluster and related quantum chemical theories for ab initio-based materials modeling.

A. Grüneis (✉)
Institute for Theoretical Physics, Vienna University of Technology, Wien, Austria
e-mail: andreas.grueneis@tuwien.ac.at

# 1      Introduction

The solution of the many-electron Schrödinger equation is at the heart of ab initio calculations in computational materials science. Owing to the complex interplay of interelectronic interactions as well as the Pauli exclusion principle for fermionic systems, the many-electron Schrödinger equation is a "non-polynomial hard" problem. Density functional theory (DFT) in the framework of approximate exchange and correlation energy functionals has revolutionized computational materials science by mapping the many-electron Schrödinger equation to a computationally tractable Kohn-Sham Hamiltonian for noninteracting electrons in an effective potential, achieving reasonable accuracy for quantum mechanical systems. However, despite the great successes of DFT during the last decades, a number of challenges in the field of electronic structure theory calculations remain unresolved. Some of the most prominent examples where currently available approximate density functionals fail in achieving a qualitatively correct description of materials properties include systems with strong electronic correlation effects, systems where nonlocal van der Waals interactions play an important role, and systems with localized electronic states that are incorrectly described due to self-interaction errors. From the perspective of quantum chemical wavefunction theories, these challenges could be dealt with by employing an appropriate ansatz for the true many-electron wavefunction, albeit increasing the computational complexity substantially.

# 2      Quantum Chemical Wavefunction Theories

In traditional quantum chemistry, the many-electron wavefunction is expanded into a basis of Slater determinants that are constructed from a set of orbitals. By construction, this basis is antisymmetric under exchange of two coordinates, as required for any fermionic wavefunction, solving the problem of self-interaction even on the lowest level of quantum chemical wavefunction-based theory, the Hartree–Fock (HF) approximation. In HF theory the many-electron wavefunction is approximated using a single Slater determinant only such that:

$$|\Phi^{\mathrm{HF}}\rangle = \frac{1}{\sqrt{N!}} \left| \begin{pmatrix} \phi_1(\mathbf{r}_1) & .. & \phi_n(\mathbf{r}_1) \\ .. & .. & .. \\ \phi_1(\mathbf{r}_n) & .. & \phi_n(\mathbf{r}_n) \end{pmatrix} \right| . \tag{1}$$

The HF determinant is constructed from a set of orthonormal one-electron orbitals, $\phi_i(\mathbf{r})$, which are obtained by minimizing the Hartree–Fock energy, whereby all the coupling terms of the Hamiltonian between the Hartree–Fock determinant and the corresponding single-excited Slater determinants vanish, which is also referred to as Brillouin's theorem. In periodic systems the index $i$ is a compound index of the Bloch wave vector $k_i$ used to sample the first Brillouin zone and the band index $n_i$. Hartree–Fock (HF) theory can be regarded as a low-rank tensor

approximation to the many-electron wavefunction, employing an antisymmetrized outer product of one-electron orbitals to approximate the many-body wavefunction. Hence, by construction, HF theory neglects electronic correlation effects that cannot be captured using products of one-electron functions only.

Traditional quantum chemical post-HF methods, such as Møller-Plesset perturbation, coupled cluster, and full configuration interaction (FCI) theories, capture electronic correlation effects by extending the wavefunction basis with additional excited HF determinants (Helgaker et al. 2000). The FCI wavefunction reads:

$$|\Psi\rangle = c_0|\Phi^{\mathrm{HF}}\rangle + \sum_{\substack{i\in\mathrm{occ.}\\ a\in\mathrm{unocc.}}} c_i^a|\Phi_i^a\rangle + \sum_{\substack{i,j\in\mathrm{occ.}\\ a,b\in\mathrm{unocc.}}} c_{ij}^{ab}|\Phi_{ij}^{ab}\rangle$$

$$+\ldots+ \sum_{\substack{i_1,..,i_n\in\mathrm{occ.}\\ a_1,..,a_n\in\mathrm{unocc.}}} c_{i_1..i_n}^{a_1..a_n}|\Phi_{i_1..i_n}^{a_1..a_n}\rangle. \tag{2}$$

$|\Phi_i^a\rangle$ correspond to singly excited determinants that are constructed from the HF determinant by replacing the $i$-th occupied orbital with the $a$-th unoccupied orbital. Higher excited determinants are obtained in an analogue manner. In $n$-electron systems we can construct at most $n$-fold excited determinants. These determinants form a complete and orthonormal basis that can be systematically improved upon toward a formally complete basis set. Computationally, the basis of (excited) Slater determinants is very convenient. All required occupied and unoccupied orbitals are obtained in the underlying HF ground state calculation. Furthermore this choice of the basis introduces a large degree of sparsity to the full many-electron Hamiltonian. Based on the Slater-Condon rules, electronic structure codes need to compute one- and two-electron integrals only that can be kept in storage or rapidly calculated on the fly. Furthermore we note that the Coulomb interaction does not couple determinants with different total momenta, increasing the sparsity of the Hamiltonian for systems with periodic boundary conditions even further. However, the downside of this approach is that the number of wavefunction coefficients in the Slater determinant basis ($c_i^a$, $c_{ij}^{ab}$, .. , $c_{i_1..i_n}^{a_1..a_n}$) grows combinatorially with respect to the number of electrons and the number of unoccupied HF orbitals. In ab initio calculations the number of unoccupied orbitals is finite only because the one-electron basis set is truncated. The latter typically corresponds to plane waves or atom-centered basis functions such as Gaussian-type orbitals that are used to expand the occupied and unoccupied one-electron orbitals. In full configuration interaction theory, no further approximations are made to the wavefunction coefficients, and the exact diagonalization of the many-electron Hamiltonian in the Slater determinant basis yields the corresponding coefficients of ground and excited many-electron states. However, in practice full configuration interaction theory is not feasible for real materials that require the treatment of systems containing more than hundred electrons and significantly larger numbers of unoccupied orbitals. Therefore wavefunction parametrizations are needed to approximate the coefficients in a more compact manner without compromising accuracy. We note that for

weakly correlated systems, these coefficients are reasonably well approximated using independent electron pair approximations such as the second-order Møller-Plesser perturbation (MP2) theory (Møller and Plesset 1934). MP2 theory employs intermediate normalization ($c_0 = 1$) and approximates the singles ($c_i^a$) and doubles ($c_{ij}^{ab}$) coefficients as well as the correlation energy in a perturbative manner. MP2 theory accounts for pair-wise additive van der Waals interactions between atoms. We stress, however, that higher-order effects such as the screening of these interactions caused by polarization of the remaining electrons in the system are neglected. Furthermore MP2 theory is not suitable for small gap and metallic systems, where any order of truncated Møller-Plesset perturbation theory fails and would require the resummation of all electron-pair coupling terms in the so-called ring approximation to account for higher-order polarization effects such that the correlation energy in metals converges. The so-called ring approximation is included in the coupled cluster singles and doubles (CCSD) theory as well as in the widely used random-phase approximation. Coupled cluster theory uses an exponential ansatz of cluster operators for the wavefunction that reads:

$$|\Psi^{\text{CC}-n}\rangle = e^{\hat{T}}|\Phi^{\text{HF}}\rangle, \tag{3}$$

where $\hat{T} = \sum_{m=1}^{n} \hat{T}_m$ and

$$\hat{T}_m|\Phi^{\text{HF}}\rangle = \sum_{\substack{i_1,...,i_m \in \text{occ.} \\ a_1,...,a_m \in \text{unocc.}}} t_{i_1..i_m}^{a_1..a_m}|\Phi_{i_1..i_m}^{a_1..a_m}\rangle. \tag{4}$$

In coupled cluster singles and doubles theory, the cluster operator reads $\hat{T} = \hat{T}_1 + \hat{T}_2$. We stress that due to the exponential ansatz of the CCSD wavefunction, the coefficients of all $i$-fold excited Slater determinants with $i \geq 2$ are being approximated using the amplitudes of the singly and doubly excited determinants; for example, $c_{ijk}^{abc} = t_i^a t_{jk}^{bc}$. Coupled cluster theory can be regarded as a low-rank tensor approximation to the exact configuration interaction wavefunction coefficients in the Slater determinant basis. The exponential ansatz used in coupled cluster theories effectively approximates the coefficients of highly excited determinants by outer products of cluster amplitudes with a lower rank. CCSD theory is exact for two-electron systems, and higher orders of coupled cluster theory become exact for systems with the corresponding number of electrons. The cluster amplitudes $t_i^a$ and $t_{ij}^{ab}$ are obtained by solving the amplitude equations $\langle\Phi_i^a|e^{-\hat{T}}He^{\hat{T}}|\Phi^{\text{HF}}\rangle = 0$ and $\langle\Phi_{ij}^{ab}|e^{-\hat{T}}He^{\hat{T}}|\Phi^{\text{HF}}\rangle = 0$, respectively. In this manner the coupling between different electron pairs is accounted for in an approximate manner including perturbation theory diagrams of a certain type to infinite order. The coupled-cluster method was initially proposed by Fritz Coester and Hermann Kümmel for applications in the field of nuclear physics (Coester 1958; Coester and Kümmel 1960). In 1966, Jiri Cizek reformulated the method for electron correlation in atoms and molecules (Cizek 1966), where it became a standard for quantum chemical

calculations on systems that do not exhibit strong static correlation (Bartlett and Musiał 2007). For most applications a good trade-off between computational cost and high accuracy is achieved by truncating the cluster operator at doubles (CCSD) and accounting for the effect of triples in a perturbative manner. This approach is referred to as CCSD(T) theory (Raghavachari et al. 1989) and achieves chemical accuracy for the prediction of reaction energies and barrier heights for a wide range of chemical reactions (Bartlett and Musiał 2007; Helgaker et al. 2000). As a consequence of the approximations to the many-electron wavefunction employed in CC theory, the treatment of strong correlation problems is extremely limited. However, this is an area of ongoing research, and the discussion of quantum chemical theories that are capable of treating strong correlation problems would be beyond the scope of this article.

One advantage of quantum chemical theories is that they constitute a hierarchy, which, starting from the one-particle Hartree–Fock (HF) approximation, allows for a systematic treatment of the quantum many-body effects that are captured with an increasing level accuracy by employing MP2, CCSD, and CCSD(T) theories. Calculated ground state properties such as cohesive energies, lattice constants, or bulk moduli of solids typically exhibit decreasing errors using higher levels of theory. Some of the most widely used ab initio codes for the study of periodic systems that have implementations of theories mentioned above include CRYSCOR, CP2K, FHI-AIMS, PYSCF, SIESTA, and VASP. We note that these implementations differ in their choice of the employed one-electron basis set and algorithms used to compute the electronic correlation energies. For the study of materials using quantum chemical theories, the calculated properties need to be carefully checked for convergence with respect to the employed computational parameters. This has to be done in a similar manner as for the more widely used density functional theory calculations. However, in practice the convergence with respect to the employed basis sets and cutoff parameters is slower than for their DFT counterparts. This is caused by the fact the wavefunction theories account for all electronic correlation effects explicitly.

## 3 Approaching the Complete Basis Set Limit

The many-electron wavefunctions introduced above are expanded in a basis of Slater determinants constructed from (unoccupied) Hartree–Fock orbitals. The computational complexity of canonical Møller-Plesset perturbation theory, coupled-cluster methods, and full configuration interaction theory scales between quadratically and exponentially with respect to the number of unoccupied orbitals. Therefore the ability to span the relevant parts of the Hilbert space with as few orbitals as possible is crucial for the implementation of efficient periodic correlated methods. In practice all calculated quantities suffer from a basis set incompleteness error that is caused by the truncation of the employed unoccupied orbital manifold. The optimal choice of the unoccupied orbital manifold minimizes the incompleteness error of the calculated quantity in a controllable manner.

Some of the most widely used basis sets for the expansion of unoccupied orbitals include plane waves and GTOs. As an illustration of their respective characteristic properties, we consider two limiting cases, the uniform electron gas and an atom in a box. From the perspective of the uniform electron gas, plane waves are the natural choice of basis to expand one- and many-electron wavefunction quantities. Plane waves are eigenfunctions of the kinetic energy operator and exhibit the same periodicity as the simulation cell. In realistic ab initio calculations, these plane waves also have a number of appealing features. A single cutoff parameter that limits the kinetic energy of the included plane waves is used to systematically expand the plane wave basis to completeness which is free from basis set superposition errors (BSSE) and linear dependencies. However, there are obvious drawbacks to plane wave expansions. They lack reference to the nature of the atomic environment, having equal basis coverage throughout the cell. This can lead to a substantial waste of computational effort when studying an atom or molecule in a box (Grüneis et al. 2011).

For atoms or molecules, GTOs form a very compact orbital basis. Their widespread use in the field of quantum chemistry has led to standardized tabulated basis sets of increasing size and flexibility (Dunning 1989; Balabanov and Peterson 2005). Orbitals beyond the core and valence shells are included to account for appropriate polarization of the atomic wavefunctions in bonding environments and to provide a description of correlation effects. Basis sets are commonly arranged in hierarchies so that they can be systematically expanded to allow for consistent and extrapolatable convergence. Gaussian-type orbitals are used in a range of periodic electronic structure codes (Dovesi et al. 2014; Pisani et al. 2012; Maschio et al. 2010; Hirata et al. 2004; Arnim and Ahlrichs 1998; Lippert et al. 1999; Krack and Parrinello 2000; Kudin and Scuseria 2000; Füsti-Molnár and Pulay 2002; VandeVondele and Hutter 2003; Sun et al. 2018). However, the introduction of such local basis sets also leads to several shortcomings such as basis set superposition errors (BSSE) and linear dependencies of diffuse atom-centered basis functions in densely packed solids. These problems can partly be accounted for by counterpoise BSSE corrections and removing linearly dependent basis functions. The local nature of these functions is often used for reduced scaling techniques in order to approach linear scaling mean-field treatments (Goedecker 1999; Strain et al. 1996; Burant et al. 1996; Goedecker and Scuseria 2003) and can also be extended to local treatment of correlation (Maschio et al. 2010; Pisani et al. 2008).

For the calculation of energy differences such as the adsorption energy of a molecule on a surface, it is beneficial to employ basis sets that can be truncated such that a large fraction of the incompleteness error cancels in a controllable manner. GTOs exhibit this advantageous property, allowing for obtaining accurate estimates of interaction energies between weakly interacting fragments such as binding energies of physisorbed molecules on surfaces, despite suffering from large incompleteness errors in the respective absolute energies. Plane wave basis sets can also be used to expand (pseudized) GTOs, inheriting these advantageous properties (Booth et al. 2016).

Another approach to obtain compact unoccupied orbital manifolds for the expansion of many-electron wavefunctions is provided by natural orbitals (Löwdin 1955).

Natural orbitals are obtained by diagonalizing the unoccupied-unoccupied orbital block of the reduced density matrix and truncating the obtained natural orbital manifold according to their occupation number. This procedure yields for many applications an optimal unoccupied orbital manifold. To reduce the computational cost of this procedure, it is possible to approximate the reduced density matrix at a lower level of theory such as direct MP2 only. Natural orbitals "downfold" the unoccupied orbitals calculated using plane wave basis sets for atoms and molecules in a box to manifolds that are similarly compact as GTOs (Grüneis et al. 2011).

Despite all the considerations outlined above, the convergence of the many-electron wavefunction and that of calculated expectation values such as the correlation energy is frustratingly slow with respect to the number of unoccupied orbitals (Helgaker et al. 2000; Hättig et al. 2012). Therefore extrapolation techniques that remove the remaining basis set incompleteness error are needed on top of these fairly large basis set calculations. In the case of plane wave basis set calculations, analytic and numerical results from MP2 theory suggest a $1/M$ decay of the basis set incompleteness error where $M$ is the number of plane waves used in the calculation, allowing for straightforward extrapolation to the CBS limit (Shepherd et al. 2012; Marsman et al. 2009). Similar scaling laws are employed for the extrapolation of correlation energies to the complete basis set limit using GTOs (Helgaker et al. 2000).

## 3.1   Explicitly Correlated Methods

The slow convergence of properties calculated using wavefunction-based methods with respect to the number of orbitals originates from the difficulty to describe the many-electron wavefunction in the vicinity of the electron cusp. As the electrons coalesce, a derivative discontinuity or "cusp" must arise, so that a divergence in the kinetic energy operator cancels an opposite one in the potential. The shape of the wavefunction at the cusp is exactly defined to first-order in the interelectronic distance by the Kato cusp conditions (Pack and Byers Brown 1966; Kato 1957). The a priori inclusion of the cusp conditions in the wavefunction ansatz is a cornerstone of explicitly correlated or so-called F12 theories (Kutzelnigg and Klopper 1991; Ten-no 2004a, b; Werner et al. 2007; Hättig et al. 2010, 2012; Grüneis et al. 2017). Explicitly correlated methods augment the conventional wavefunction expansions discussed in the previous section with additional terms that account for the cusp conditions explicitly. Since electronic correlation is for the most part a short-ranged phenomenon, the proper description of the wavefunction shape at short interelectronic distances allows for capturing the largest fraction of the correlation energy in solids and molecules.

Figure 1a depicts the different contributions to the explicitly correlated first-order wavefunction for two electrons in a box with a homogeneous background. It can be seen that the resultant F12 wavefunction exhibits a cusp at the electron coalescence point. Within an expansion of Slater determinants, the exact cusp is never obtained, and a quantitatively correct linear form at small interelectronic
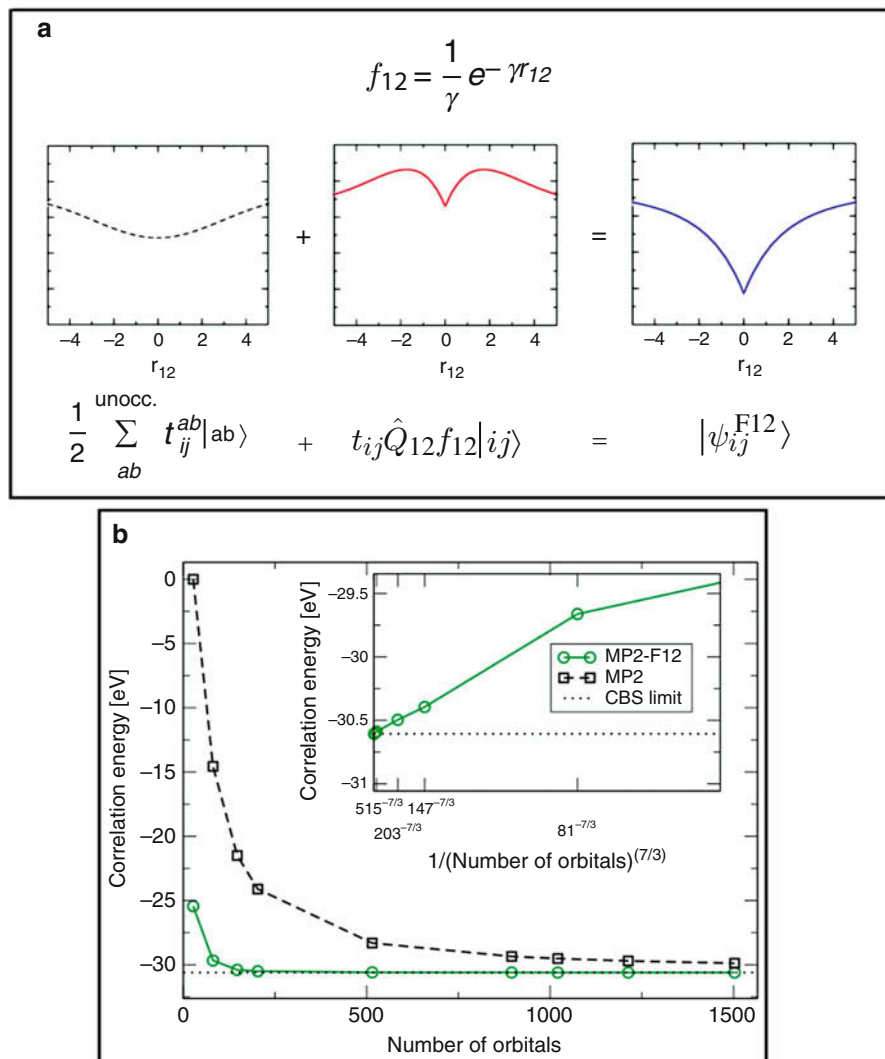
**Fig. 1** (**a**) Illustration of the different contributions to the explicitly correlated two-electron wavefunction in a box with a homogeneous background. $f_{12}$ is the Slater-type correlation factor with the variational parameter $\gamma$. (**b**) Convergence of the MP2 and MP2-F12 correlation energies for the 54 electron UEG simulation cell ($r_s = 5.0$ bohr) with respect to the employed number of orbitals $M$ (Grüneis et al. 2013)

distances only arises with large basis sets of high momenta. Explicitly correlated methods provide a more compact ansatz for the many-electron wavefunction by augmenting a traditional Slater determinant expansion with geminals that are centered at the electron coalescence points, satisfying the first-order cusp condition

exactly. The first-order pair functions are expanded as:

$$|u_{ij}^{F12}\rangle = \frac{1}{2}\sum_{ab} t_{ij}^{ab}|ab\rangle + t_{ij}\hat{Q}_{12}f_{12}|ij\rangle. \tag{5}$$

The first and second term on the right-hand side of the above equation correspond to the conventional MP1 pair function and the strongly orthogonal geminals, respectively. $t_{ij}$ are geminal amplitudes determined by the universal cusp conditions, and $f_{12}$ is the correlation factor that models the shape of the correlation hole and is typically chosen to be a Slater-type function. This choice ensures that the geminal functions included in the basis are linear with respect to $r_{12}$ in the vicinity of the electron-electron cusp and decay to zero at large $r_{12}$, where the wavefunction is expected to vary smoothly and is generally well-represented by the conventional determinantal basis. The projector $\hat{Q}_{12}$ enforces strong orthogonality between the first-order and the HF wavefunction, and it also enforces orthogonality between the standard and F12 contributions to the first-order wavefunction. Figure 1b shows that the MP2-F12 correlation energy converges much more rapidly than the MP2 correlation energy with respect to the employed basis set ($\approx \frac{1}{M^2}$), reducing the required basis set significantly without compromising accuracy. Even though the findings shown in Fig. 1b were obtained for homogeneous systems, similar basis set reductions can be achieved for ab initio systems. F12 theories introduce the need for three- and sometimes even four-electron integrals. The direct calculation of these integrals is circumvented in practice by the introduction of complementary auxiliary basis sets, significantly increasing the prefactor of the computational complexity for F12 calculations (Klopper and Samson 2002). As a result explicitly correlated methods are only advantageous for more complex parent methods such as CC theories (Knizia et al. 2009; Hättig et al. 2010). In recent years these developments are also being adapted to the field of periodic correlated methods, showing great promise in expanding the scope of quantum chemical wavefunction theories for solids (Usvyat 2013; Grüneis 2015).

## 4 Approaching the Thermodynamic Limit

Quantum chemical wavefunction-based methods are most widely used for the study of atoms and molecules. In contrast to molecular systems, properties of solids and surfaces need to be calculated in the thermodynamic limit (TDL) in order to allow for a direct comparison to experiment. The thermodynamic limit is approached as $N \rightarrow \infty$, where $N$ is the number of particles in the simulation (super-)cell while the density is kept constant. The TDL is usually approached by sampling the first Brillouin zone using increasingly dense $k$-meshes. However, the convergence of calculated properties to the thermodynamic limit is very slow, often exceeding the computational resources of even modern supercomputers due to the steep scaling of the computational complexity of most post-HF methods with respect to system

size. We stress that many properties such as the binding energy of molecules on surfaces converge slower than their counterparts calculated on the level of mean-field theories such as DFT. This originates from the fact that correlated post-HF methods capture long-range electronic correlation effects such as van der Waals interactions explicitly. Even though the corresponding long-ranged contribution to the electronic correlation energy is small compared to short-ranged correlation energy contributions, the accumulation of weak van der Waals interactions can become a non-negligible contribution to the property of interest. Different strategies have been developed to correct for finite size errors that are defined as the difference between the thermodynamic limit and the finite simulation cell results. Local theories that employ correlation energy expressions depending on localized electron pairs can approximate correlation energy contributions of long-distant pairs using computationally more efficient yet less accurate theories. Alternatively local theories can account for electron pairs that are disregarded based on a distance criterion by correcting using an $R^{-6}$-type extrapolation (Usvyat et al. 2012). Canonical implementations of periodic post-HF methods employ scaling laws for extrapolations to the TDL that are based on an analogue rationale (Booth et al. 2013; Ben et al. 2013; McClain et al. 2017). We note in passing that auxiliary field quantum Monte Carlo theory employs finite size corrections that are based on parametrized density functionals obtained from finite uniform electron gas simulation cells (Kwee et al. 2008).

The problem of slow thermodynamic limit convergence and concomitantly large finite size errors is a common feature of quantum Monte Carlo (QMC) and correlated quantum chemical methods. QMC methods employ correction schemes that reduce finite size errors originating from the evaluation of the kinetic and potential energy operators in the many-electron Hamiltonian (Chiesa et al. 2006; Kwee et al. 2008; Holzmann et al. 2016). The most notable corrections include structure factor interpolation techniques and twist averaging. The latter reduces finite size errors originating from the one-electron operators, whereas the structure factor interpolation method allows for reducing finite size errors originating from the two-electron operator. Recently we have introduced finite size correction methods that are inspired by the corresponding techniques used in QMC calculations. These finite size corrections allow for achieving thermodynamic limit results of solids and surfaces using quantum chemical wavefunction theories in a very efficient manner (Gruber et al. 2018; Liao and Grüneis 2016), reducing the computational cost significantly.

## 5    Reduced-Scaling Approximations

In this last section, we briefly mention promising developments from the field of quantum chemical wavefunction-based methods that allow for reducing the scaling of the computational complexity. The steep polynomial scaling of the computational complexity with respect to system size of canonical CCSD ($\mathcal{O}(N^6)$) or CCSD(T) ($\mathcal{O}(N^7)$) theories arises from the use of spatially delocalized canonical

orbitals. Canonical orbitals are conceptually and computationally convenient. They are orthogonal and diagonalize the Fock matrix, greatly simplifying the post-HF correlation schemes. However, these orbitals are spatially delocalized, and their use does not allow one to exploit the fact that electronic correlation is an intrinsically short-ranged phenomenon. The use of spatially localized, instead of canonical, orbitals allows to construct MP2 and CC algorithms that scale more favorably with system size, down to even $\mathcal{O}(N)$, at the price of a significant increase in complexity of the underlying equations with respect to their canonical counterparts. The most notable among them are (based on) the local correlation method of Pulay and Saebø (Pulay and Saebø 1986; Saebø and Pulay 1993; Schütz et al. 1999; Subotnik and Head-Gordon 2005), the so-called local ansatz of Stollhof and Fulde (Stollhoff and Fulde 1978, 1980; Stollhoff 1996), the method of increments of Stoll (Stoll 1992a, b, c; Paulus 2006; Schwerdtfeger et al. 2010; Rościszewski et al. 1999), the atomic orbital basis formulations of Ayala and Scuseria (Ayala and Scuseria 1999; Scuseria and Ayala 1999), or the use of truncated pair natural orbitals (Neese et al. 2009; Kubas et al. 2016; Ma et al. 2017). Other recent developments to reduce the scaling of the computational cost with respect to system size include techniques that calculate the MP2 correlation energy using stochastic methods (Willow et al. 2012; Neuhauser et al. 2013). Furthermore we note that the change of the representation also allows for reducing the scaling of the computational cost without necessarily requiring a local orbital representation (Mardirossian et al. 2018; Schäfer et al. 2017; Friesner 1985).

Finally we note that tensor rank decomposition techniques and so-called tensor hypercontraction methods have also successfully lead to a reduction of the computational cost. The tensor hypercontraction (THC) technique introduced by Hohenstein et al. (Hohenstein et al. 2012a, b; Parrish et al. 2012) performs a low-rank tensor decomposition of the Coulomb integrals. Furthermore, Shenvi et al. (2013) used a similar approach. In the work of Benedikt et al., it was shown that tensor rank decomposition techniques allow for a low-rank decomposition of the coupled cluster amplitudes directly (Benedikt et al. 2013).

The above discussion of reduced scaling approximations is by no means complete but reflects the effort of the electronic structure theory community to reduce the computational cost and expand the scope of quantum chemical methods for the study of materials further.

## 6    Conclusion

In conclusion, we have summarized recent developments in the field of quantum chemical wavefunction-based theories that show great promise of expanding the scope of quantum chemical wavefunction theories for materials modeling. A significant amount of research effort is expended on accelerating the convergence of the many-electron wavefunction approximations and resultant expectation values with respect to the employed orbital basis set, system size, and other computational parameters. Furthermore local approximations allow for the reduction of the scaling

of the computational complexity with respect to system size, making studies of materials containing more than hundred atoms possible. These developments pave the way for a routine use of accurate quantum chemical theories such as coupled cluster methods in the field of surface chemistry and solid-state physics. The ability to predict accurate benchmark results will help the entire electronic structure theory community to improve further upon computationally more efficient yet less accurate ab initio workhorse theories.

# References

Arnim MV, Ahlrichs R (1998) Parallelization of density functional and RI-coulomb approximation in turbomole. J Comput Chem 19:1746

Ayala PY, Scuseria GE (1999) Linear scaling second-order Møller–Plesset theory in the atomic orbital basis for large molecular systems. J Chem Phys 110(8):3660–3671. http://scitation.aip.org/content/aip/journal/jcp/110/8/10.1063/1.478256

Balabanov N, Peterson K (2005) Systematically convergent basis sets for transition metals. I. All-electron correlation consistent basis sets for the 3D elements Sc–Zn. J Chem Phys 123:064107

Bartlett RJ, Musiał M (2007) Coupled-cluster theory in quantum chemistry. Rev Mod Phys 79:291–352. https://link.aps.org/doi/10.1103/RevModPhys.79.291

Ben MD, Hutter J, Vandevondele J (2013) Electron correlation in the condensed phase from a resolution of identity approach based on the Gaussian and Plane Waves scheme Electron correlation in the condensed phase from a resolution of identity approach based on the Gaussian and Plane Waves scheme. J Chem Theory Comput 9(6):2654–2671. https://doi.org/10.1021/ct4002202

Benedikt U, Böhm KH, Auer AA (2013) Tensor decomposition in post-Hartree–Fock methods. II. CCD implementation. J Chem Phys 139(22):224101. http://scitation.aip.org/content/aip/journal/jcp/139/22/10.1063/1.4833565

Booth GH, Grüneis A, Kresse G, Alavi A (2013) Towards an exact description of electronic wavefunctions in real solids. Nature 493(7432):365–370. https://doi.org/10.1038/nature11770, http://www.ncbi.nlm.nih.gov/pubmed/23254929

Booth GH, Tsatsoulis T, Chan GKL, Grüneis A (2016) From plane waves to local Gaussians for the simulation of correlated periodic systems. J Chem Phys 145(8):084111. http://scitation.aip.org/content/aip/journal/jcp/145/8/10.1063/1.4961301

Burant JC, Scuseria GE, Frisch MJ (1996) A linear scaling method for hartree-fock exchange calculations of large molecules. J Chem Phys 105:8969–8972

Chiesa S, Ceperley DM, Martin RM, Holzmann M (2006) Finite-size error in many-body simulations with long-range interactions. Phys Rev Lett 97(7):6–9. https://doi.org/10.1103/PhysRevLett.97.076404

Cizek J (1966) On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell–type expansion using quantum–field theoretical methods. J Chem Phys 45(11):4256–4266. https://doi.org/10.1063/1.1727484

Coester F (1958) Bound states of a many-particle system. Nucl Phys 1:421–424

Coester F, Kümmel H (1960) Short-range correlations in nuclear wave functions. Nucl Phys 17:477–485

Dovesi R, Orlando R, Erba A, Zicovich-Wilson CM, Civalleri B, Casassa S, Maschio L, Ferrabone M, De La Pierre M, D'Arco P, Noel Y, Causa M, Rerat M, Kirtman B (2014) Crystal14: a program for the ab initio investigation of crystalline solids. Int J Quantum Chem 114: 1287–1317

Dunning TH (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J Chem Phys 90:1007–1023

Friesner RA (1985) Solution of self-consistent field electronic structure equations by a pseudospectral method. Chem Phys Lett 116(1):39–43. https://doi.org/10.1016/0009-2614(85)80121-4, http://www.sciencedirect.com/science/article/pii/0009261485801214

Füsti-Molnár L, Pulay P (2002) The fourier transform coulomb method: efficient and accurate calculation of the coulomb operator in a gaussian basis. J Chem Phys 117:7827

Goedecker S (1999) Linear scaling electronic structure methods. Rev Mod Phys 71:1085–1123

Goedecker S, Scuseria GE (2003) Linear scaling electronic structure methods in chemistry and physics. Comput Sci Eng 5:14–21

Gruber T, Liao K, Tsatsoulis T, Hummel F, Grüneis A (2018) Applying the coupled-cluster ansatz to solids and surfaces in the thermodynamic limit. Phys Rev X 8:021043. https://link.aps.org/doi/10.1103/PhysRevX.8.021043

Grüneis A (2015) Efficient explicitly correlated many-electron perturbation theory for solids: application to the schottky defect in MgO. Phys Rev Lett 115:066402. https://link.aps.org/doi/10.1103/PhysRevLett.115.066402

Grüneis A, Booth GH, Marsman M, Spencer J, Alavi A, Kresse G (2011) Natural orbitals for wave function based correlated calculations using a plane wave basis set. J Chem Theory Comput 7(9):2780–2785. https://doi.org/10.1021/ct200263g

Grüneis A, Shepherd JJ, Alavi A, Tew DP, Booth GH (2013) Explicitly correlated plane waves: accelerating convergence in periodic wavefunction expansions. J Chem Phys 139(8):084112

Grüneis A, Hirata S, Ohnishi Yy, Ten-no S (2017) Perspective: explicitly correlated electronic structure theory for complex systems. J Chem Phys 146(8):080901. https://doi.org/10.1063/1.4976974

Hättig C, Tew DP, Köhn A (2010) Communications: accurate and efficient approximations to explicitly correlated coupled-cluster singles and doubles, CCSD-F12. J Chem Phys 132(23):231102

Hättig C, Klopper W, Köhn A, Tew DP (2012) Explicitly correlated electrons in molecules. Chem Rev 112(1):4–74

Helgaker T, Jørgensen P, Olsen J (2000) Molecular electronic-structure theory. Wiley, New York

Hirata S, Podeszwa R, Tobita M, Bartlett RJ (2004) Coupled-cluster singles and doubles for extended systems. J Chem Phys 120(6):2581–2592. https://doi.org/10.1063/1.1637577

Hohenstein EG, Parrish RM, Martínez TJ (2012a) Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller–Plesset perturbation theory. J Chem Phys 137(4):044103. http://scitation.aip.org/content/aip/journal/jcp/137/4/10.1063/1.4732310

Hohenstein EG, Parrish RM, Sherrill CD, Martínez TJ (2012b) Communication: tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions. J Chem Phys 137(22):221101. http://scitation.aip.org/content/aip/journal/jcp/137/22/10.1063/1.4768241

Holzmann M, Clay RC, Morales MA, Tubman NM, Ceperley DM, Pierleoni C (2016) Theory of finite size effects for electronic quantum monte carlo calculations of liquids and solids. Phys Rev B 94:035126. https://link.aps.org/doi/10.1103/PhysRevB.94.035126

Kato T (1957) On the eigenfunctions of many-particle systems in quantum mechanics. Commun Pure Appl Math 10(2):151–177

Klopper W, Samson CCM (2002) Explicitly correlated second-order Møller–Plesset methods with auxiliary basis sets. J Chem Phys 116(15):6397–6410

Knizia G, Adler TB, Werner HJ (2009) Simplified CCSD(T)-F12 methods: theory and benchmarks. J Chem Phys 130(5):054104

Krack M, Parrinello M (2000) All-electron ab-initio molecular dynamics. Phys Chem Chem Phys 2:2105

Kubas A, Berger D, Oberhofer H, Maganas D, Reuter K, Neese F (2016) Surface adsorption energetics studied with "gold standard" wave-function-based ab initio methods: Small-molecule binding to tio2(110). J Phys Chem Lett 7(20):4207–4212. https://doi.org/10.1021/acs.jpclett.6b01845

Kudin KN, Scuseria GE (2000) Linear-scaling density-functional theory with gaussian orbitals and periodic boundary conditions: efficient evaluation of energy and forces via the fast multipole method. Phys Rev B 61:16440

Kutzelnigg W, Klopper W (1991) Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory. J Chem Phys 94(3):1985–2001

Kwee H, Zhang S, Krakauer H (2008) Finite-size correction in many-body electronic structure calculations. Phys Rev Lett 100(12):126404. http://link.aps.org/doi/10.1103/PhysRevLett.100.126404

Liao K, Grüneis A (2016) Communication: finite size correction in periodic coupled cluster theory calculations of solids. J Chem Phys 145(14):0–4. https://doi.org/10.1063/1.4964307

Lippert G, Hutter J, Parrinello M (1999) The gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. Theory Chem Acc 103:124

Löwdin PO (1955) Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. Phys Rev 97:1474–1489. https://link.aps.org/doi/10.1103/PhysRev.97.1474

Ma Q, Schwilk M, Köppl C, Werner HJ (2017) Scalable electron correlation methods. 4. Parallel explicitly correlated local coupled cluster with pair natural orbitals (pno-lccsd-f12). J Chem Theory Comput 13(10):4871–4896. https://doi.org/10.1021/acs.jctc.7b00799, pMID:28898081

Mardirossian N, McClain JD, Chan GKL (2018) Lowering of the complexity of quantum chemistry methods by choice of representation. J Chem Phys 148(4):044106. https://doi.org/10.1063/1.5007779

Marsman M, Grüneis A, Paier J, Kresse G (2009) Second-order Møller–Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set. J Chem Phys 130(18):184103

Maschio L, Usvyat D, Schütz M, Civalleri B (2010) Periodic local Møller–Plesset second order perturbation theory method applied to molecular crystals: study of solid $NH_3$ and $CO_2$ using extended basis sets. J Chem Phys 132:134706

McClain J, Sun Q, Chan GKL, Berkelbach TC (2017) Gaussian-based coupled-cluster theory for the ground state and band structure of solids. J Chem Theory Comput 13(3):1209–1218. https://doi.org/10.1021/acs.jctc.7b00049, http://arxiv.org/abs/1701.04832

Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46(7):618–622. http://link.aps.org/doi/10.1103/PhysRev.46.618

Neese F, Wennmohs F, Hansen A (2009) Efficient and accurate local approximations to coupled-electron pair approaches: an attempt to revive the pair natural orbital method. J Chem Phys 130(11):114108. http://scitation.aip.org/content/aip/journal/jcp/130/11/10.1063/1.3086717

Neuhauser D, Rabani E, Baer R (2013) Expeditious stochastic approach for MP2 energies in large electronic systems. J Chem Theory Comput 9(1):24–27

Pack RT, Byers Brown W (1966) Cusp conditions for molecular wavefunctions. J Chem Phys 45(2):556–559

Parrish RM, Hohenstein EG, Martínez TJ, Sherrill CD (2012) Tensor hypercontraction. II. Least-squares renormalization. J Chem Phys 137(22):224106. http://scitation.aip.org/content/aip/journal/jcp/137/22/10.1063/1.4768233

Paulus B (2006) The method of increments—a wavefunction-based ab initio correlation method for solids. Phys Rep 428(1):1–52. https://doi.org/10.1016/j.physrep.2006.01.003, http://www.sciencedirect.com/science/article/pii/S0370157306000330

Pisani C, Maschio L, Casassa S, Halo M, Schütz M, Usvyat D (2008) Periodic local MP2 method for the study of electronic correlation in crystals: theory and preliminary applications. J Comput Chem 29:2113–2124

Pisani C, Schütz M, Casassa S, Usvyat D, Maschio L, Lorenz M, Erba A (2012) Cryscor: a program for the post-hartree-fock treatment of periodic systems. Phys Chem Chem Phys 14:7615

Pulay P, Saebø S (1986) Orbital-invariant formulation and second-order gradient evaluation in Møller–Plesset perturbation theory. Theor Chim Acta 69(5):357–368. https://doi.org/10.1007/BF00526697

Raghavachari K, Trucks GW, Pople JA, Head-Gordon M (1989) A fifth-order perturbation comparison of electron correlation theories. Chem Phys Lett 157(6):479–483. https://doi.org/10.1016/S0009-2614(89)87395-6, http://www.sciencedirect.com/science/article/pii/S0009261489873956

Rościszewski K, Paulus B, Fulde P, Stoll H (1999) Ab initio calculation of ground-state properties of rare-gas crystals. Phys Rev B 60:7905–7910. http://link.aps.org/doi/10.1103/PhysRevB.60.7905

Saebø S, Pulay P (1993) Local treatment of electron correlation. Ann Rev Phys Chem 44(1):213–236. https://doi.org/10.1146/annurev.pc.44.100193.001241

Schäfer T, Ramberger B, Kresse G (2017) Quartic scaling MP2 for solids: a highly parallelized algorithm in the plane wave basis. J Chem Phys 146(10):104101. https://doi.org/10.1063/1.4976937

Schütz M, Hetzer G, Werner HJ (1999) Low-order scaling local electron correlation methods. I. Linear scaling local MP2. J Chem Phys 111(13):5691–5705. https://doi.org/10.1063/1.479957

Schwerdtfeger P, Assadollahzadeh B, Hermann A (2010) Convergence of the Møller–Plesset perturbation series for the FCC lattices of neon and argon. Phys Rev B Condens Matter 82(20):205111

Scuseria GE, Ayala PY (1999) Linear scaling coupled cluster and perturbation theories in the atomic orbital basis. J Chem Phys 111(18):8330–8343. http://scitation.aip.org/content/aip/journal/jcp/111/18/10.1063/1.480174

Shenvi N, Aggelen Hv, Yang Y, Yang W, Schwerdtfeger C, Mazziotti D (2013) The tensor hypercontracted parametric reduced density matrix algorithm: coupled-cluster accuracy with O(r4) scaling. J Chem Phys 139(5):054110. http://scitation.aip.org/content/aip/journal/jcp/139/5/10.1063/1.4817184

Shepherd JJ, Grüneis A, Booth GH, Kresse G, Alavi A (2012) Convergence of many-body wave-function expansions using a plane-wave basis: from homogeneous electron gas to solid state systems. Phys Rev B Condens Matter 86(3):035111

Stoll H (1992a) The correlation energy of crystalline silicon. Chem Phys Lett 191(6):548–552. https://doi.org/10.1016/0009-2614(92)85587-Z, http://www.sciencedirect.com/science/article/pii/000926149285587Z

Stoll H (1992b) Correlation energy of diamond. Phys Rev B 46:6700–6704. http://link.aps.org/doi/10.1103/PhysRevB.46.6700

Stoll H (1992c) On the correlation energy of graphite. J Chem Phys 97(11):8449–8454. http://scitation.aip.org/content/aip/journal/jcp/97/11/10.1063/1.463415

Stollhoff G (1996) The local ansatz extended. J Chem Phys 105(1):227–234. http://scitation.aip.org/content/aip/journal/jcp/105/1/10.1063/1.471867

Stollhoff G, Fulde P (1978) Description of intraatomic correlations by the local approach. Zeitschrift für Physik B Condensed Matter 29(3):231–237. https://doi.org/10.1007/BF01321187

Stollhoff G, Fulde P (1980) On the computation of electronic correlation energies within the local approach. J Chem Phys 73(9):4548–4561. http://scitation.aip.org/content/aip/journal/jcp/73/9/10.1063/1.440693

Strain MC, Scuseria GE, Frisch MJ (1996) Achieving linear scaling for the electronic quantum Coulomb problem. Science 271:51–53

Subotnik JE, Head-Gordon M (2005) A local correlation model that yields intrinsically smooth potential-energy surfaces. J Chem Phys 123(6):064108. https://doi.org/10.1063/1.2000252

Sun Q, Berkelbach TC, Blunt NS, Booth GH, Guo S, Li Z, Liu J, McClain J, Sayfutyarova ER, Sharma S, Wouters S, Chan GK-L (2018) PySCF: the python–based simulations of chemistry framework. Wiley Interdiscip Rev Comput Mol Sci 8(1):e1340. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340

Ten-no S (2004a) Explicitly correlated second order perturbation theory: introduction of a rational generator and numerical quadratures. J Chem Phys 121(1):117–129. https://doi.org/10.1063/1.1757439

Ten-no S (2004b) Initiation of explicitly correlated slater-type geminal theory. Chem Phys Lett 398(1):56–61. https://doi.org/10.1016/j.cplett.2004.09.041, http://www.sciencedirect.com/science/article/pii/S000926140401379X

Usvyat D (2013) Linear-scaling explicitly correlated treatment of solids: periodic local MP2-F12 method. J Chem Phys 139(19):194101

Usvyat D, Sadeghian K, Maschio L, Schütz M (2012) Geometrical frustration of an argon monolayer adsorbed on the MgO (100) surface: an accurate periodic ab initio study. Phys Rev B 86:045412. https://link.aps.org/doi/10.1103/PhysRevB.86.045412

VandeVondele J, Hutter J (2003) An efficient orbital transformation method for electronic structure calculations. J Chem Phys 118:4365

Werner HJ, Adler TB, Manby FR (2007) General orbital invariant MP2-F12 theory. J Chem Phys 126(16):164102

Willow SY, Kim KS, Hirata S (2012) Stochastic evaluation of second-order many-body perturbation energies. J Chem Phys 137(20):204122. https://doi.org/10.1063/1.4768697

# Optimal Control Theory for Electronic Structure Methods

# 22

Alberto Castro

## Contents

### Abstract

Optimal control theory (OCT) is a branch of mathematics that deals with the problem of finding optimal trajectories for dynamical systems. It can be used in combination with time-dependent quantum mechanical methods that describe the evolution of the electronic and/or nuclear wave functions of atoms, molecules, or materials in the presence of external perturbations, such as electromagnetic fields. OCT may then find the optimal shape of those external perturbations: the optimal character is defined in terms of a functional of the behavior of the system. This chapter provides a brief description of the basic elements of the theory and an overview of its applications to quantum dynamics and electronic structure.

A. Castro (✉)
ARAID Foundation and Institute for Biocomputation and Physics of Complex Systems of the University of Zaragoza, Zaragoza, Spain
e-mail: acastro@bifi.es

# 1 Introduction

*Spectroscopy* studies the interaction of radiation with matter. Its parenthood is attributed to Newton, who explained the famous prism experiments that decomposed light into its colored components. He used for the first time the word *spectrum* (that means *appearance*, *image*, or *apparition*) to refer to the decomposed light image ["Comparing the length of this coloured spectrum with its breadth (. . . )" (Newton 1671)], initiating the field of research and giving it its name at the same time.

The essence of the typical spectroscopic experiment has not changed: take some form of radiation, make it interact with a piece of matter, and observe the radiation reflected or transmitted by it. The resulting data, typically radiation intensities as a function of their wavelength, have continued to be called *spectra* since Newton. By interpreting these spectra, we learn how matter reacts to the perturbation, and, most importantly, we can deduce its microscopic structure. Spectroscopy is the key to understanding how matter looks like at the scale of its constituents.

Moreover, if radiation induces a reaction, it means that the state of matter can be manipulated or *controlled* by tuning the radiation characteristics: expose skin to light, and you can control its tan by adjusting exposition time, modulating its intensity with sunscreen, etc. Initially, only *incoherent control* could be exercised, since all available sources emitted incoherent radiation. One could only manipulate the state of macroscopic portions of matter by tuning time-averaged characteristics of the radiation, since the many photons of an incoherent source impinge stochastically on the microscopic elements of the sample.

This state of affairs changed with the invention of the laser (Maiman 1960): the photons could then be emitted coherently, the frequency could be tuned to the point of almost monochromaticity, and very high intensities could be obtained. The laser is the ideal tool for the manipulation of matter at its microscopic scale. *Coherent control* (Rice and Zhao 2000; Brumer and Shapiro 2003), also known as *quantum control*, was born. In essence, it is some kind of *inverse spectroscopy*: the goal is not to understand how a piece of matter (atom, molecule, nanostructure, solid material, etc.) reacts to an external electromagnetic field but, inversely, to find out what is the precise form of an external field that may induce a given reaction.

The first attempts targeted selective photochemistry, in particular the goal of the cleavage of molecules by some particular bond, with a simple concept: one uses the remarkable monochromaticity that a laser source can achieve to tune the light to the characteristic frequency of that bond. If enough energy can be deposited on it, the bond will break. This can only work if a particular frequency can be associated to the given bond (vibrational molecular modes may not correspond to single-bond vibrations but may represent collective motions). However, the biggest problem is intramolecular vibrational redistribution (IVR) (Bloembergen and Zewail 1984): the energy quickly flows to other degrees of freedom, before the wanted reaction takes place. The final outcome is an increase in the molecular temperature – as it happens when incoherent irradiation is used – and perhaps its fracture at unexpected places.

In the following section, the early proposals of coherent control that overcame this limitation are briefly reviewed. However, the great leap forward was achieved thanks to the improvement of *pulse shapers* – capable of creating almost arbitrary pulses – and the *adaptive feedback control* proposal (AFC) by Judson and Rabitz (1992). This will also be described in the next section. Optimal control theory (OCT) (Kirk 1998) is presented in Sect. 3; its application in combination with electronic structure theories will finally be discussed in Sect. 4.

## 2   Early Coherent Control Proposals and the Adaptive Feedback Control Setup

Given the inability of single monochromatic laser beams to control the complexity of chemical reactions and other quantum processes, the next obvious step is the use of two frequencies. Brumer and Shapiro (1986a, b, 1989) analyzed the problem of unimolecular reactions triggered by light (i.e., photodissociation) and proposed the control through the quantum interference of two pathways: suppose that two reaction channels are accessible from a given initial state and the energy necessary for it can be supplied by two different paths, one single photon of a given frequency and a multiple number of photons of smaller (commensurate) frequency. Furthermore, suppose that the two beams can act *coherently* on the system. Then, the relative phase of the two beams can be adjusted, and the yields of the various channels will depend on the interference between them. The modification of the relative phase (as well as, possibly, the relative intensity) becomes in this way a control mechanism.

A different approach was taken by Tannor and Rice (1985) and Tannor et al. (1986), who proposed the use of femtosecond pulses in a *pump-dump* setup, illustrated in Fig. 1. A compound "ABC" is initially in its ground-state configuration; two dissociative channels, $AB + C$ and $A + BC$, exist in the ground-state potential energy surface but are adiabatically blocked by a barrier. An initial *pump* pulse at $t_0$ can be used to place the system in an electronic excited state surface (or perhaps in a linear combination of those). The system then evolves, and a second *dump* pulse can be used at different times, with different effects: if applied at $t_1$, the system is transferred back to the electronic ground state, but on the dissociative channel $AB + C$; if applied later at $t_2$, the system will dissociate to $A + BC$.

There have been other "early" quantum control scheme proposals – such as the stimulated Raman adiabatic passage (STIRAP) (Gaubatz et al. 1988; Kuklinski et al. 1989), etc. However, all those first schemes were characterized by the use of relatively simple field shapes and few control parameters (the relative phases, the time delay, etc.). It soon became obvious that they could all be regarded as particular forms of a more general concept: the use of arbitrary laser pulse shapes, with durations and frequencies of the order of the mechanism that is to be controlled, and the variation of many control parameters, controlling those shapes.
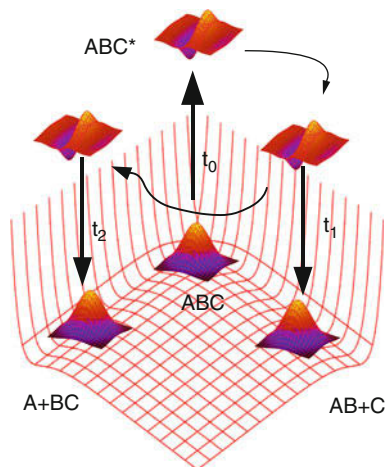
**Fig. 1** The pump-dump control setup proposed by Tannor, Kosloff, and Rice (Tannor and Rice 1985; Tannor et al. 1986). A compound in its initial "ABC" ground-state configuration is *pumped* at $t_0$ with a femtosecond laser pulse to an electronic excited state. It then evolves freely until a later *dump* pulse transfers it back to the ground state potential energy surface; depending on the delay time, $t_1$ or $t_2$, the system may land on different dissociative channels, $AB + C$ or $A + BC$

The key instrument to realize this ambitious concept is the laser shaper, a device capable of designing the temporal shape of a pulse. This tool quickly evolved during the 1990s (Weiner 2000). Essentially, it works in the following way: A grating decomposes a simply shaped laser pulse into its frequency components, each of which passes through a filter, capable of modifying its intensity, phase, or both. Then, a second grating is used to combine all the components into the output, *shaped* pulse.

Based on the laser shapers, Judson and Rabitz (1992) proposed the adaptive feedback control (AFC) setup. It is illustrated in Fig. 2. Suppose that we wish to find a laser pulse capable of producing some given reaction. This fact can be measured, for example, by using a time-of-flight mass spectrometer. The failure or success of the pulse can be summarized by a *performance function*, defined in terms of the experiment outcome. The idea then is to create a *learning loop*: the value of the performance function is fed to an optimization algorithm, which decides on the value of the parameters that determine the shape of the next laser pulse to be probed. This loop is iterated a number of times, until an optimal pulse is found.

The AFC mechanism has become the most successful control technique for quantum processes (see Brif et al. (2010) for a rather exhaustive list of applications). Its success is based on various facts: (1) the remarkable capacity that laser shapers have acquired to produce almost arbitrary pulses with high intensities, (2) the high repetition rate that the AFC loop normally has, (3) the existence of efficient optimization algorithms for many-variable functions, and (4) the robustness of the solutions found by the AFC technique.
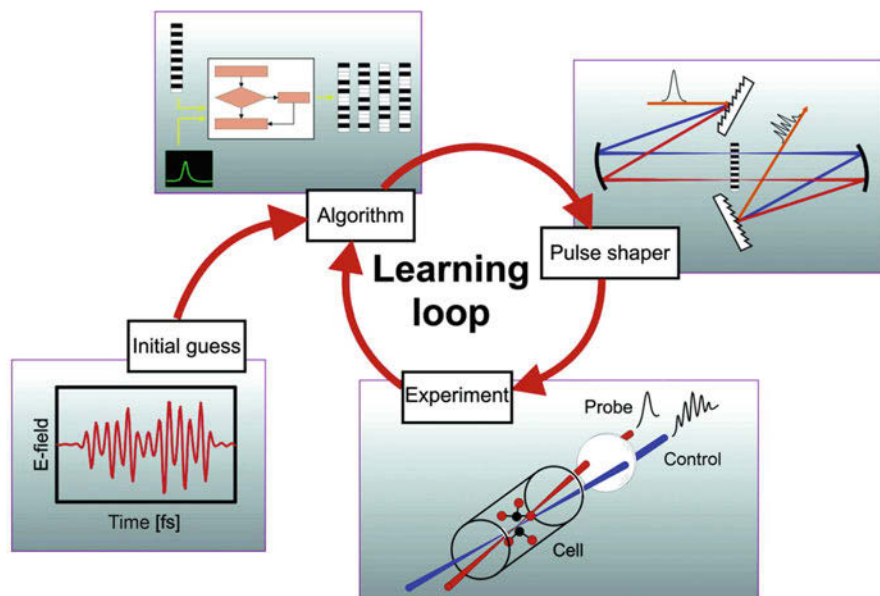
**Fig. 2** The adaptive feedback control (AFC) experimental setup (Judson and Rabitz 1992). (From Rabitz et al. (2000). Reprinted with permission from AAAS)

## 3   Optimal Control Theory: Mathematical Framework

In parallel to the experimental advances described above, a complementary theoretical framework was necessary to describe and solve the problem of quantum control, i.e., finding the parameters that define some external action on some model for a quantum system, capable of forcing it to behave in some target manner. This description of the problem falls within the scope of control theory, a branch of engineering and mathematics that was however invented for other problems. Optimal control theory (OCT) (Kirk 1998) is a particularly successful form of control theory. This section is a crash course on OCT. Its application to quantum problems (quantum OCT, i.e., QOCT), with focus on electronic systems, will be the topic of the next section.

The state of a system, within the mathematical framework of dynamical systems, is typically modeled as a point $y$ in some smooth manifold; this state is fully described by a set of coordinates $y_1, \ldots, y_n \equiv y$ (assuming the manifold dimension to be finite). The system evolves according to a *dynamical law*:

$$\dot{y}(t) = f(y(t), u(t), t). \tag{1}$$

$$y(t_0) = y_0. \tag{2}$$

The functions $u(t) = u_1(t), \ldots, u_M(t)$ are the *control inputs* or *control functions*. The idea behind the definition of the *state* of a system is the following: it is a set of numbers (that change in time, as the state changes) such that, given their knowledge at a given time $t_0$, fully determine the behavior of the system at times $t \geq t_0$ (given fixed control inputs). If the evolution of the system is smooth in time, this means that the equation of motion or dynamical law is a first-order differential equation such as Eq. (1).

In a typical formulation of a control problem, the evolution of the system is usually constrained to a set of *admissible* trajectories that we may denote by $Y$. For example, one may require that the starting state is given: $y \in Y \Rightarrow y(t_0) = y_0$. Likewise, the control inputs may be constrained in multiple ways, and therefore we must define a set of admissible controls $U$.

Sometimes, the control inputs are not presented as functions but as real numbers, usually called the *control parameters*. Note that there is no loss of generality in considering one or the other option: in any numerical implementation of the problem, each function is represented by a set of parameters; inversely, any set of parameters can be considered as a set of functions that are constrained to be constant in time.

The previous Eqs. (1) and (2) are not general enough to describe any possible physical model: stochastic terms, for example, are not included, which exclude relevant models such as Langevin's equations. Of course, extensions of the basic framework of OCT exist for more general situations. However, many important processes fall into this category, such as any quantum process described by Schrödinger's equation. In this case, the dynamical law is linear:

$$f(y(t), u(t), t) = -i H[u(t), t] y(t), \tag{3}$$

where $H[u(t), t]$ is the Hamiltonian, a linear Hermitian operator.

In the realm of electronic structure theory, the state $y$ is a many-electron wave function. However, more often than not, one does not solve the full many-electron problem, too large for more than a few electrons, but uses some of the many electronic structure methods that have been developed over the years, for example, multi-configuration time-dependent Hartree-Fock (Nest et al. 2005; Beck et al. 2000) or time-dependent density-functional theory (Marques et al. 2012), which both have been combined with OCT (Mundt and Tannor 2009; Castro et al. 2012).

If the control inputs $u$ are given, the evolution of the system is fully determined by its equations of motion. However, if we can *control* the values of these parameters, then we can manipulate the evolution of the system. This is the gist of control theory; "optimal" control theory, in particular, pursues this control by defining a *performance measure* (also known sometimes as *target functional*), a function dependent on the system trajectory, and searching for the control parameters that maximize it. In some cases, the problem statement obviously and uniquely determines the definition for the performance function, whereas in some other cases, many possible valid options exist, and the practitioner must choose one. A very generic form for this function is:

$$F(y, u) = h(y(T)) + \int_{t_0}^{T} dt \; g(y(t), u(t), t) \,. \tag{4}$$

This definition separates a *terminal* part $h(y(T))$, which depends on the final system state, and a *time-dependent* or *trajectory* part, $\int_{t_0}^{T} dt \; g(y(t), u(t), t)$, which depends on the full evolution of the system. The time $T$ is the final propagation time. This may be fixed or not: an important class of control problems is the minimization of the time that it takes for some process to happen.

We already have all the necessary ingredients to formally formulate the *optimal control problem*: find an admissible control $u^* \in U$, such that it forces the system to follow an admissible trajectory $y^* \in Y$:

$$\dot{y}^*(t) = f(y^*(t), u^*(t), t) \,, \tag{5}$$

$$y^*(t_0) = y_0^* \,, \tag{6}$$

and such that it maximizes the value of the performance measure, i.e.:

$$F(y^*, u^*) = h(y^*(T)) + \int_{t_0}^{T} dt \; g(y^*(t), u^*(t), t) \geq F(y, u) \tag{7}$$

for any other admissible $y$ and $u$. Note that, since the specification of $u$ determines the system evolution, i.e., $u \to y[u]$, one may reformulate this as the maximization of a function of the control inputs alone:

$$G(u) = F(y[u], u) \,. \tag{8}$$

Also note that, one may indistinctly speak of minimization or maximization; the formalism is identical except for a sign in the performance function (that typically changes its name to *cost function* when the goal is its minimization).

One very common example of control problem in the quantum world is the following: suppose that we wish to find the precise shape of the amplitude of an electric field, $u(t)$, such that it has *moderate intensity*, and it maximizes the population of some excited states $y_I$ at the final propagation time $T$. Then, a possible definition for the performance measure is:

$$F(y, u) = |\langle y(T)|y_I\rangle|^2 - \alpha \int_{0}^{T} dt \; u^2(t) \,. \tag{9}$$

The second term somehow codifies the requirement of moderate intensity and is generally called a *penalty* term. Of course, multiple other definitions are possible (given the imprecise character of the *moderate intensity* requirement), and this fact exemplifies the lack of uniqueness in the definition for the performance function that control problems often have.

There are two main mathematical approaches to the control theory problem: the dynamic programming method developed by Bellman (1957) and the method based on Pontryagin's minimum principle (Boltyanskii et al. 1956; Pontryagin et al. 1962). Both are briefly described in the following two subsections.

## 3.1    Dynamic Programming

This approach is based on the *principle of optimality*:

*If a trajectory $y^*(t)$, corresponding to the control inputs $u^*(t)$, is optimal for a given performance measure defined in the interval $[t_0, T]$, then any final piece of that trajectory (i.e., the restriction of $y^*(t)$ to an interval $[t_a, T]$ for any $t_a \in (t_0, T)$), with the same control inputs, is also optimal for the same performance measure.*

This is in fact a rather simple principle, and it can be very easily proved by reductio ad absurdum (Kirk 1998). It can be used to solve the control problem in two different equivalent ways:

- Frequently, the problem is first discretized in time, and the set of admissible states is also fully enumerated (it may happen that the problem is already discrete from the beginning and is posed as a series of decisions). Then, the principle of optimality can be used to create a recurrence relation that fully determines the optimal solution. One starts from the last part of the trajectory, whose optimal solution must be a part of the full solution according to the principle of optimality. Then, a step-by-step backward repetition of this idea is used.
- One may directly derive a partial differential equation that constitutes a necessary and sufficient condition for a set of control inputs and corresponding trajectories to be optimal: the Hamilton-Jacobi-Bellman (HJB) equation. Of course, its solution then requires an a posteriori discretization.

The HJB equation, or the equivalent discretized recurrence algorithms, apparently supplies a complete solution to the optimal control problem. Unfortunately, it is not always useful due to the bad scaling of the computational cost with the problem variables: number of time steps, system state dimensions, and number of control inputs. This drawback was already identified by Bellman, who named it the *curse of dimensionality*, and it is probably the reason behind the fact that dynamic programming has been rarely used for quantum control problems.

## 3.2    Pontryagin's Minimum Principle

Pontryagin et al. (Boltyanskii et al. 1956) established a set of necessary conditions for the optimal control inputs and state trajectory of a control problem, which is known as Pontryagin's minimum principle. This is not the place to present it in detail, or to demonstrate it, but it is worth to outline a simplified version.

Given a control problem defined by the dynamical system, Eqs. (1) and (2), and by the performance function defined in Eq. (4), we define a *Hamiltonian function* (not to be mistaken with the Hamiltonian of the system, if it exists):

$$\mathcal{H}(y(t), u(t), p(t), t) = g(y(t), u(t), t) + p^T(t) \cdot f(a(y(t), u(t), t)). \quad (10)$$

This definition needs a new set of variables, $p(t) = p_1(t), \ldots, p_n(t)$, sometimes called the *costate*, or the momentum variables. The following are necessary conditions for $u^*(t)$ and $y^*(t)$ to minimize the performance functions:

$$\dot{y}^*(t) = \nabla_p \mathcal{H}(y^*(t), u^*(t), p^*(t), t), \quad (11)$$

$$y^*(t_0) = y_0, \quad (12)$$

$$\dot{p}^*(t) = -\nabla_y \mathcal{H}(y^*(t), u^*(t), p^*(t), t), \quad (13)$$

$$p^*(T) = \nabla_y h(y(T)), \quad (14)$$

$$\mathcal{H}(y^*(t), u^*(t), p^*(t), t) \leq \mathcal{H}(y^*(t), u(t), p^*(t), t) \text{ for any } u \in U. \quad (15)$$

Note that this is a simplified version of the principle that assumes that the final time $T$ is fixed, the initial state is also fixed, and, in contrast, the final state at time $T$ is not constrained in any way. However, it is sufficient to display one of its key traits: the use of an auxiliary set of variables, $p(t)$, that verify equations of motion that are similar, but not identical, to the ones verified by the original variables. The boundary conditions specified in Eq. (14), in particular, are given at the final time $T$, which means that the computation of $p$ must be obtained via a backward propagation.

## 4    Application to Electron Dynamics

Pontryagin's minimum principle is in fact at the theoretical base of most applications of OCT to quantum problems. Or, put differently, the equations used in the field of QOCT can be considered re-derivations of Pontryagin's minimum principle for particular cases, even if it is often uncredited. For example, a prototypical QOCT problem may be defined by the target in Eq. (9) and by an evolution governed by Schrödinger's equation – i.e., the linear relation in Eq. (3). Let us assume furthermore that the Hamiltonian has the form:

$$H[u(t), t] = H_0 + u(t)D. \quad (16)$$

$H_0$ may be the field-free Hamiltonian of a molecule, whereas $D$ may be its dipole coupling to the electric field component of a laser pulse whose amplitude is $u(t)$ (assuming the dipole approximation). Then, one can prove that the optimal $u^*$ is given by:

$$u^*(t) = -\frac{1}{\alpha}\text{Im}\langle p^*(t)|D|y^*(t)\rangle\,, \tag{17}$$

where $y^*(t)$ and the auxiliary wave function $p^*(t)$ are fully determined by:

$$\dot{y}^*(t) = -i\,H[u^*(t), t]y^*(t)\,. \tag{18}$$

$$y^*(T) = y_0\,, \tag{19}$$

$$\dot{p}^*(t) = -i\,H[u^*(t), t]p^*(t)\,. \tag{20}$$

$$p^*(T) = \langle y_I|y^*(T)\rangle y_I\,. \tag{21}$$

These equations can be derived by application of Pontryagin's principle or from scratch by making use of variational calculus.

Equation (17) is in fact the zero condition for the variation of the performance function with respect to the control input. Specifically,

$$\frac{\delta G}{\delta u(t)} = 2\text{Im}\langle p[u](t)|D|y[u](t)\rangle - 2\alpha u(t)\,, \tag{22}$$

where, remember, $G(u) = F(y[u], u)$ [Eq. (8)]. Setting the previous functional derivative to zero is obviously a requisite [i.e., Eq. (17)] for the control input (the laser pulse, in our example) to be optimal.

Many algorithms aimed at finding the solution control given in Eq. (17) have been developed over the years. A very successful one was designed by Zhu and Rabitz (1998). We describe it in Fig. 3 as an example, especially since it is the scheme used

---

1. Choose an initial guess solution $u^{(0)}(t)$.
2. Propagate $y_0 \rightarrow y(T)$ using the Hamiltonian $H_0 + u^{(0)}(t)$. The solution is $y^{(0)}$.
3. Let $p^{(0)}(T) = \langle y_I|y^{(0)}(T)\rangle y_I$.
4. Propagate $p^{(0)}(T) \rightarrow p^{(0)}(0)$ using the Hamiltonian $H_0 + u^{(0)}(t)$. The solution is $p^{(0)}$.
5. Let $u^{(1)}(t) = -\frac{1}{\alpha}\text{Im}\langle p^0(t)|D|y^0(t)\rangle$.
6. Let $k = 1$.
7. Propagate $y_0 \rightarrow y(T)$ using the Hamiltonian $H_0 + u^{(k)}(t)$. The solution is $y^{(k)}$.
8. Let $p^{(k)}(T) = \langle y_I|y^{(k)}(T)\rangle y_I$.
9. Propagate $p^{(k)}(T) \rightarrow p^{(k)}(0)$ using the Hamiltonian $H_0 + u^{(k)}(t)$. The solution is $p^{(k)}$.
10. Let $u^{(k+1)}(t) = -\frac{1}{\alpha}\text{Im}\langle p^k(t)|D|y^k(t)\rangle$.
11. If $u^{(k+1)} \approx u^{(k)}$, stop and $u^* = u^{(k+1)}$ is the solution. Otherwise, repeat steps 7–11 until convergence.

**Fig. 3**   The monotonic algorithm of Zhu and Rabitz (1998)

for one of the examples below. Some of the most common features of all quantum control algorithms can be learned from the analysis of this one, namely:

- The algorithm requires backward propagations of the auxiliary state $p$ (since the boundary condition is established at the final propagation time $T$). This is not surprising, as this fact is already implicit in Pontryagin's minimum principle. Note, however, that not all schemes require backward propagations: they are only needed if the scheme makes use of the gradient of the target functional with respect to the control parameters – or the functional derivative with respect to $u(t)$ if a continuous function is used as control. Many algorithms only require function evaluations (e.g., evolutionary algorithms), and therefore the only operation required is the propagation of the equations of motion.
- Convergence to the global optimum is not guaranteed. The algorithm of Zhu and Rabitz is always monotonic, but in general, no algorithm can guarantee that a global optimum will be found. Multiple local solutions may exist.
- The cost of the optimization grows linearly with the cost of propagating the wave functions, as this is the basic operation of the algorithm. The number of propagations needed to reach convergence depends however on the scheme and grows with the number of control parameters or functions.

The previous discussion, based on Eqs. (16), (17), (18), (19), (20), (21), and (22), considers a continuous function of time, $u(t)$. Therefore, the "gradient" of the functional [Eq. (22)] is in fact a functional derivative. It is worth to reconsider the theory assuming that the control inputs are a set of parameters $u_1, \ldots, u_M$ – especially since, upon discretization for their numerical representation, the control input functions are also reduced to a finite set of parameters. Schrödinger's equation would then read:

$$\dot{y}(t) = -iH[u, t]y(t),  \tag{23}$$

$$y(t_0) = y_0.  \tag{24}$$

and the performance function can be, for example:

$$F(y, u) = \langle y(T)|O|y(T)\rangle - P(u),  \tag{25}$$

where $O$ is some operator whose expectation value at the end of the process we wish to maximize (the case $O = |y_I\rangle\langle y_I|$ corresponds with the excited state population described above) and $P(u)$ is a penalty function (which, as in the example above, may be added to avoid unwanted features of the control parameters, such as too high intensities). The previous functional derivative is now a normal gradient, given by:

$$\frac{\partial G}{\partial u_k}(u) = 2\text{Im}\int_{t_0}^{T} dt \ \langle p[u](t)|\frac{\partial H}{\partial u_k}(u, t)|y[u](t)\rangle - \frac{\partial P}{\partial u_k}(u).  \tag{26}$$

where the auxiliary state $p[u]$ is defined by its equations of motion:

$$\dot{p}[u](t) = -iH[u,t]p[u](t)\,, \tag{27}$$

$$p[u](T) = Oy[u](T)\,. \tag{28}$$

Note the remarkable similarity of these equations to the ones that define the state: $p$ also follows Schrödinger's equation, and the only difference is the initial value condition that changes to a final value one. This similarity is a consequence of (1) the linear character of Schrödinger's equation, (2) the quadratic character of the target functional with respect to the state, and (3) the fact that the target functional only depends on the state at the final time of the propagation, $T$. If any of these conditions is not met, the equations of motion for $p$ become more complex, including nonlinear or inhomogeneous terms.

These QOCT equations, or variations of it, have been applied many times in numerous investigations, since the first attempts in the 1980s (Shi et al. 1988; Peirce et al. 1988; Shi and Rabitz 1989; Kosloff et al. 1989). Most of these, up to recently, have focused on the control, with femtosecond pulses, of the motion of the nuclear wave packet on one or a few potential energy surfaces. That motion typically happens on a time scale of hundreds of femtoseconds or picoseconds. However, laser technology has improved dramatically, reaching the attosecond time scale when the millennium turned (Krausz and Ivanov 2009). This access to the attosecond scale has enabled the possibility of not only observing electron dynamics in their natural time scale but also controlling them.

Therefore, the combination of QOCT with non-equilibrium quantum dynamics methods has become necessary to match the experimental advances. Unfortunately, modeling electron dynamics is notoriously difficult (van Leeuwen and Stefanucci 2013). The main reason is the exponential growth of computational cost and memory requirements with the number of electrons involved in the calculations. For this reason, the simple Schrödinger equation is substituted with more elaborate models that operate on smaller objects, at the cost of losing the structural simplicity of the original linear equation. This problem becomes especially relevant for QOCT calculations: the expression for the gradient of the performance function with respect to the control parameters involves the forward propagation of the system state, and the backward propagation of a fictitious *costate*. This means that performing QOCT calculations is as costly as performing a number of propagations for the system and model employed.

In the following, in order to show a glimpse of the possibilities that the combination of QOCT with electronic dynamics methods offers, three sample cases are briefly presented.

## 4.1    One Electron Case: Control of Motion in Quantum Rings

If there is only one electron (or if the electron-electron interaction can be neglected), electron dynamics is actually easy and manageable, and quantum optimal control
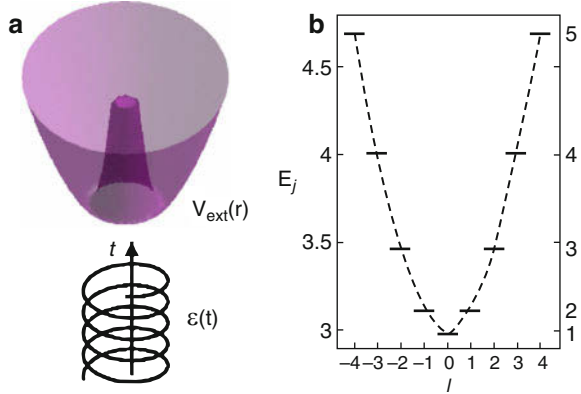
can be performed by making use of the very same equations described above for the case in which the system simply follows Schrödinger's equation.

For example, Räsänen and collaborators (Räsänen et al. 2007) studied single electrons in quantum rings: the rings are in fact two-dimensional traps built in semiconductor heterostructures, where the electrons of the two-dimensional electron gas formed at the interface can be confined. They can be modeled with simple Hamiltonians, i.e.:

$$H_0 = -\frac{\hbar^2}{2m^*}\nabla^2 + \frac{1}{2}m^*\omega_0^2 r^2 + V_0 e^{-r^2/d^2} . \tag{29}$$

This is a 2D Hamiltonian ($r^2 = x^2 + y^2$), whose confinement potential is depicted in Fig. 4a: one may see how the potential well is ring-shaped. Note that to describe these systems, one makes use of the effective mass approximation: the parameter $m^*$ is an effective mass (in this particular example, it was set to 0.067 times the real electron mass, the experimental value measured in GaAs).

The energy levels of this Hamiltonian are also depicted in Fig. 4b. The $x$ axis represents the orbital angular momentum: positive and negative values correspond to clockwise and anticlockwise electron rotation within the ring. The goal of this study was to fully control the rotation of the electron, by placing the electron into any of those excited states, making use of two-component electric fields, i.e., the time-dependent Hamiltonian is given by:

$$H(t) = H_0 - x\varepsilon_x(t) - y\varepsilon_y(t). \tag{30}$$

The fields $\varepsilon_x$ and $\varepsilon_y$ are the control inputs. The optimization was achieved by making use of the rapidly convergent algorithm described above in Fig. 3 (Zhu and Rabitz 1998). The target or performance function was set to be, as above, the population of the excited states, in order to design electric pulses capable of changing the electron rotation.

The use of QOCT in this work led to finding optimal shapes for the controlling fields that are more complex than simple continuous wave pulses tuned to the energy differences of the states. Those optimal pulses could arbitrarily transfer the population between states almost at will in very short times.

## 4.2    QOCT in Combination with Multicomponent Hartree-Fock

In order to go beyond single-electron cases and attempt the control of larger electronic systems, one must make use of one of the many approaches to the time-dependent many-electron problem that have been developed since the dawn of quantum mechanics. Mundt and Tannor were the first to attempt the combination of one of these schemes, multi-configuration time-dependent Hartree-Fock (MCT-DHF) (Nest et al. 2005; Beck et al. 2000), with QOCT (Mundt and Tannor 2009).

In MCTDHF, the very large and unmanageable $N$-electron wave function $\Psi(x_1, \ldots, x_N, t)$ (where $x_i$ represents the position and spin coordinates of one electron) is substituted by the smaller ansatz:

$$\Psi^{MCTDHF}(x_1, \ldots, x_N, t) = \sum_{k_1}^{K} \cdots \sum_{k_N}^{K} c_{k_1, \ldots, k_N}(t) \Pi_{l=1}^{N} \varphi_{k_l}(x_l, t). \tag{31}$$

The single-particle orbitals $\{\varphi_k\}$ define a basis in the single-particle Hilbert space and may change in time along with the coefficients $c_{k_1, \ldots, k_N}(t)$. The number of orbitals $K$ determines the accuracy of the method. The evolution of this new wave function – that means the evolution of the coefficients and of the orbitals – is determined by the Dirac-Frenkel variational principle (Raab 2000), which ensures the minimum possible distance to the true solution. The resulting equations are rather involved, and it is not worth to repeat them here. The key point, however, is that it is no longer a simple set of linear equations such as Schrödinger's. As a consequence, the control equations also become very involved, and, in fact, they were not derived and used in the work of Mundt and Tannor. In contrast, the idea was to use the simple control expressions valid for Schrödinger's equation and use the MCTDHF wave function as an approximation to the true wave function.

This approach can only be valid if the MCTDHF is of very high quality: the number of orbitals of the single-electron basis is very large, in which case the wave function tends to the exact one. Unfortunately, this also means that the computational cost must be very large, too: the cost grows with $\binom{K}{N}$. And, in any case, the obtained optimal control inputs will be sub-optimal, due to the fact that the true control equations for the model are not used. Due to these reasons, this approach has not been widely followed yet.

## 4.3     QOCT in Combination with Time-Dependent Density-Functional Theory

A different approach to the time-dependent many-electron problem is time-dependent density-functional theory (TDDFT) (Marques et al. 2012). Density-based methods substitute the many-electron wave function by the one-electron density as the working variable, an idea that in the case of TDDFT is founded on the one-to-one correspondence between densities and external potentials proved in Runge and Gross (1984). In fact, one can further prove (van Leeuwen 1999) that the time-dependent electron density of a system can be reproduced by a unique external potential in a system with a different electron-electron interaction. If the two systems are identical (i.e., both have the same electron-electron interaction), the uniqueness implies the one-to-one correspondence mentioned above. And, if the first system is the real one but the second one is a noninteracting system, the theorem shows that one can find an external potential for this fictitious system, which reproduces the density of the real one (the *Kohn-Sham (KS) potential*).

This noninteracting system can be modeled with a set of one-particle equations, the time-dependent Kohn-Sham (TDKS) equations:

$$i\frac{\partial}{\partial t}\varphi_i(\mathbf{r}, t) = -\frac{1}{2}\nabla^2\varphi_i(\mathbf{r}, t) + v_{KS}[n](\mathbf{r}, t)\varphi_i(\mathbf{r}, t) \quad (i = 1, \ldots, N/2), \quad (32)$$

$$n(\mathbf{r}, t) = 2\sum_{i=1}^{N/2} |\varphi_i(\mathbf{r}, t)|^2. \quad (33)$$

The orbitals $\varphi_i$ form a single Slater determinant (we assume them here to be doubly occupied, with one pair of spin-up and one spin-down electrons occupying each spatial orbital), which is enough for noninteracting electrons. $n(\mathbf{r}, t)$ is both the real and the fictitious time-dependent electron density. The KS potential $v_{KS}$ is given by:

$$v_{KS}[n](\mathbf{r}, t) = v_{ext}(\mathbf{r}, u, t) + v_{H}[n](\mathbf{r}, t) + v_{xc}[n](\mathbf{r}, t), \quad (34)$$

where $v_{ext}$ is the true external potential, which must be fully determined by the specification of the set of control inputs $u \equiv u_1, \ldots, u_M$. $v_{H}[n](\mathbf{r}, t) = \int d\mathbf{r}' \frac{n(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|}$ is the electrostatic potential, and $v_{xc}[n](\mathbf{r}, t)$ is the exchange and correlation potential, whose precise form is unknown and must be approximated. In purity, at each time $t$, it depends on all density values at $t' \leq t$. Normally, however, the *adiabatic approximation* is assumed:

$$v_{xc}[n](\mathbf{r}, t) \approx v_{xc}^{GS}[n_t](\mathbf{r}), \quad (35)$$

where $v_{xc}^{GS}[n_t](\mathbf{r})$ is one of the approximations to the exchange and correlation potential functional used in ground-state DFT. The notation of Eq. (32) can be shortened with matrix notation:

$$i\dot{\underline{\varphi}}(t) = \hat{\underline{H}}[n(t), u, t]\underline{\varphi}(t) \,, \tag{36}$$

where we use the underlined symbol $\underline{\varphi}$ to denote a vector containing all KS orbitals and we use the matrix notation $\overline{H}[n(t), u, t] = H_{KS}[n(t), u, t]\underline{I}$ ($\underline{I}$ is the $N/2$-dimensional unit matrix). $H_{KS}[n(t), u, t]$ is the KS Hamiltonian that comprises all the terms in the right-hand side of Eq. (32).

Once the model is fully specified, the control equations can be derived. First, the performance function must be given:

$$F(\underline{\varphi}, u) = F^{\text{td}}(\underline{\varphi}, u) + F^{\text{term}}(\underline{\varphi}(T), u) \,. \tag{37}$$

The first term depends on the full propagation, whereas the second is the *terminal target* that only depends on the state at the end of the propagation. Considering that $u \rightarrow \underline{\varphi}[u]$ (i.e., the specification of the control inputs determines the system trajectory) and defining, once again, $G(u) = F(\underline{\varphi}[u], u)$, one can prove that

$$\nabla_u G(u) = \nabla_u \left. F(\underline{\varphi}, u)\right|_{\underline{\varphi}=\underline{\varphi}[u]} + 2\text{Im}\left[\sum_{j=1}^{N}\int_0^T dt \,\langle \chi_j[u](t)|\nabla_u v_{\text{ext}}(u, t)|\varphi_j[u](t)\rangle\right] \tag{38}$$

where the costate $\underline{\chi}$ is determined by the equations:

$$i\dot{\underline{\chi}}(t) = \left[\hat{\underline{H}}_{KS}[n[u](t), u, t] + \hat{\underline{K}}[\underline{\varphi}[u](t)]\right]\underline{\chi}(t) - i\frac{\delta F^{\text{td}}}{\delta\underline{\varphi}^*} \,, \tag{39}$$

$$\underline{\chi}(T) = \frac{\delta F^{\text{term}}}{\delta\underline{\varphi}^*(T)} \,. \tag{40}$$

The optimal solution is of course found at one of the zeroes of the gradient in Eq. (38). To note:

- The nonlinearity of the TDKS equations causes the appearance of the operator matrix $\hat{\underline{K}}[\underline{\varphi}[u](t)]$. The (operator) matrix elements of this object are given by:

$$\langle\mathbf{r}|\hat{K}_{ij}[\underline{\varphi}[u](t)]|\psi\rangle = -2i\varphi_i[u](\mathbf{r}, t)\times$$

$$\text{Im}\left[\int d^3r' \psi^*(\mathbf{r}') f_{\text{Hxc}}[n[u](t)](\mathbf{r}, \mathbf{r}')\varphi_j[u](\mathbf{r}', t)\right], \tag{41}$$

where $f_{\text{Hxc}}$ is the *kernel* of the Kohn-Sham Hamiltonian, defined as:

$$f_{\text{Hxc}}[n](\mathbf{r}, \mathbf{r}') = \frac{1}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta v_{\text{xc}}[n](\mathbf{r})}{\delta n(\mathbf{r}')} \,. \tag{42}$$

- Equation (39) is inhomogeneous, due to the presence of the functional derivative of $F^{\text{td}}$ with respect to $\underline{\varphi}$ (the last term in the right-hand side). This term disappears of course if the target functional is only of terminal type.
- The previous scheme permits to control the Kohn-Sham system. However, the goal is to control the *real* system. In principle, one would be tempted to design a target in terms of the real many-electron wave function – an object that however is not provided by TDDFT, which only provides the density. Therefore, the ideal situation would be that in which the performance function can be defined only through the density $n$. In this manner, optimizing for the Kohn-Sham system is strictly equivalent to optimizing for the real one.

This formalism has been implemented in the Octopus code (Marques et al. 2003; Castro et al. 2006). An illustrative simple example was described in Castro et al. (2012): the charge transfer of two electrons between two neighboring asymmetric potential wells, which is defined with the simple model:

$$v_0(x, y) = \frac{1}{64}x^4 - \frac{1}{4}x^2 + \frac{1}{32}x^3 + \frac{1}{2}y^2. \tag{43}$$

This potential landscape is depicted in Fig. 5a. A ground-state DFT calculation performed with the 2D LDA exchange and correlation parameterization provided in Attaccalite et al. (2002) sets the two electrons on the left well [see Fig. 5b].

The control mechanism is an electric field, polarized along the $x$ direction. Its amplitude is parameterized by its Fourier coefficients $u_j$, which constitute the control inputs, i.e.:

$$v_{\text{ext}}(\mathbf{r}, u, t) = v_0(x, y) + \sum_j u_j g_j(t)x, \tag{44}$$

where $g_j(t)$ are the Fourier basis functions (normalized sines and cosines). The coefficients are constrained to enforce $\varepsilon(0) = \varepsilon(T) = 0$. The performance function is defined as:

$$F[\underline{\varphi}, u] = \int_{x>0} d^2r \, n(\mathbf{r}, T) - \alpha \int_0^T dt \, \varepsilon^2[u](t), \tag{45}$$

an expression that encodes the goal: at the final time of the propagation ($T$), there should be as much charge as possible at the right well ($x > 0$). The last term of Eq. (45) is a penalty that prevents the solution field from having too much intensity.

The optimal field is shown in Fig. 5c. In order to find the solution, in this case, a standard conjugate gradient algorithm was used. After around 60 CG iterations (the convergence plot is also shown in Fig. 5), the calculation was converged. The optimal pulse successfully transfers the charge from the left well to the right well. The final charge density is plotted in Fig. 5b.
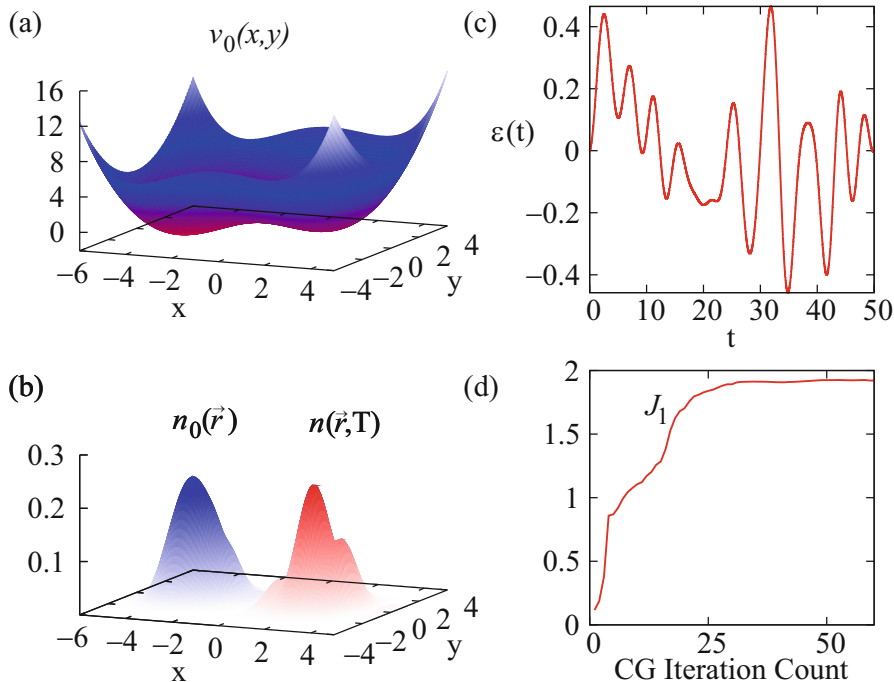
**Fig. 5** (**a**) External potential defining a model for a double quantum dot. (**b**) Density of the initial ground state [blue, $n_0(\mathbf{r})$] and final propagated density [red, $n(\mathbf{r}, T)$]. (**c**) Optimized electric field for the charge-transfer process described in the text. (**d**) Convergence history of the conjugate gradient algorithm. All magnitudes are given in effective atomic units (Reprinted with permission from Castro et al. (2006). Copyright 2012 by the American Physical Society)

Therefore, TDDFT can be combined with QOCT. This combination has been explored a few times, even sometimes in combination with nonadiabatic molecular dynamics to allow for the movement of the nuclei (Castro 2016; Walkenhorst et al. 2016; Gómez Pueyo et al. 2016). The scheme is numerically challenging but tractable. Note, however, that the use of TDDFT always implies an approximation and, in consequence, an error, which may be especially worrisome when employing high-intensity fields. This fact was analyzed, for example, in Raghunathan and Nest (2011).

## 5    Conclusions

The application of OCT to problems in the quantum realm (QOCT) has already some history and can be considered a mature theory, with numerous applications in various fields. QOCT is a generic framework, and its basic formulation makes no assumption on the nature and modeling of the quantum system to be optimized.

It must be *combined* with a given level of theory. In practice, the solution of the QOCT equations requires multiple propagations of the system under study. The cost of applying QOCT is typically therefore multiple times the cost of simulating the system once, and in consequence, it is determined by the cost of the chosen level of theory.

For many-electron systems, the propagations with truly ab initio and precise levels of theory become prohibitively expensive as the number of electrons grows, and usually few-level simplifications and models are used when attempting the combination with QOCT. These simplifications may not be accurate enough: the strong pulses that have nowadays become available imply the occupation of many excited states, and perturbative expansions also fail with increasing intensities. The advances in experimental techniques that permit the access to the attosecond time scale make it necessary to work on the possible combination of QOCT with first-principles methods. TDDFT is one possible choice.

The possibilities that these combinations of QOCT with non-equilibrium time-dependent electron dynamics methods offer are numerous: shaping of the high-harmonic-generation spectrum (i.e., quenching or increasing given harmonic orders), selective excitation of electronic excited states that are otherwise difficult to reach with conventional pulses, control of the electronic current in molecular junctions, etc.

# References

Attaccalite C, Moroni S, Gori-Giorgi P, Bachelet GB (2002) Correlation energy and spin polarization in the 2D electron gas. Phys Rev Lett 88(25):256601

Beck M, Jackle A, Worth G, Meyer HD (2000) The multiconfiguration time-dependent hartree (mctdh) method: a highly efficient algorithm for propagating wavepackets. Phys Rep 324(1): 1–105

Bellman R (1957) Dynamic programming, 1st edn. Princeton University Press, Princeton

Bloembergen N, Zewail AH (1984) Energy redistribution in isolated molecules and the question of mode-selective laser chemistry revisited. J Phys Chem 88(23):5459–5465

Boltyanskii VG, Gamkrelidze RV, Pontryagin LS (1956) On the theory of optimal processes (Russian). Dokl Akad Nauk SSSR (NS) 110:7

Brif C, Chakrabarti R, Rabitz H (2010) Control of quantum phenomena: past present and future. New J Phys 12(7):075,008

Brumer P, Shapiro M (1986a) Coherent radiative control of unimolecular reactions. Three-dimensional results. Faraday Discuss Chem Soc 82:177–185

Brumer P, Shapiro M (1986b) Control of unimolecular reactions using coherent light. Chem Phys Lett 126(6):541–546

Brumer P, Shapiro M (1989) Coherence chemistry: controlling chemical reactions [with lasers]. Acc Chem Res 22(12):407–413

Brumer P, Shapiro M (2003) Principles of the quantum control of molecular processes. Wiley, New York

Castro A (2016) Theoretical shaping of femtosecond laser pulses for molecular photodissociation with control techniques based on Ehrenfest's dynamics and time-dependent density functional theory. ChemPhysChem 17(11):1601–1607

Castro A, Appel H, Oliveira M, Rozzi CA, Andrade X, Lorenzen F, Marques MAL, Gross EKU, Rubio A (2006) Octopus: a tool for the application of time-dependent density functional theory. Phys Status Solidi (b) 243:2465–2488

Castro A, Werschnik J, Gross EKU (2012) Controlling the dynamics of many-electron systems from first principles: a combination of optimal control and time-dependent density-functional theory. Phys Rev Lett 109:153603

Gaubatz U, Rudecki P, Becker M, Schiemann S, Külz M, Bergmann K (1988) Population switching between vibrational levels in molecular beams. Chem Phys Lett 149(5):463–468

Gómez Pueyo A, Budagosky M JA, Castro A (2016) Optimal control with nonadiabatic molecular dynamics: application to the Coulomb explosion of sodium clusters. Phys Rev A 94:063421

Judson RS, Rabitz H (1992) Teaching lasers to control molecules. Phys Rev Lett 68:1500–1503

Kirk DE (1998) Optimal control theory. An introduction. Dover Publications, Inc., New York

Kosloff R, Rice SA, Gaspard P, Tersigni S, Tannor DJ (1989) Wavepacket dancing: achieving chemical selectivity by shaping light pulses. Chem Phys 139:201–220

Krausz F, Ivanov M (2009) Attosecond physics. Rev Mod Phys 81:163–234

Kuklinski JR, Gaubatz U, Hioe FT, Bergmann K (1989) Adiabatic population transfer in a three-level system driven by delayed laser pulses. Phys Rev A 40:6741–6744

Maiman TH (1960) Stimulated optical radiation in ruby. Nature 187:493–494

Marques MAL, Maitra NT, Nogueira FMS, Gross EKU, Rubio A (eds) (2012) Fundamentals of time-dependent density functional theory. Springer, Berlin/Heidelberg

Marques MAL, Castro A, Bertsch GF, Rubio A (2003) Octopus: a first-principles tool for excited electron-ion dynamics. Comput Phys Commun 151:60–78

Mundt M, Tannor DJ (2009) Optimal control of interacting particles: a multi-configuration time-dependent Hartree-Fock approach. New J Phys 11(10):105038

Nest M, Klamroth T, Saalfrank P (2005) The multiconfiguration time-dependent Hartree-Fock method for quantum chemical calculations. J Chem Phys 122(12):124102

Newton I (1671) A letter of Mr. Isaac Newton, Professor of the Mathematics in the University of Cambridge; containing his new theory about light and colors. Philos Trans 6:3075–3087

Peirce A, Dahleh M, Rabitz H (1988) Optimal control of quantum-mechanical systems: existence, numerical approximation, and applications. Phys Rev A 37(12):4950

Pontryagin LS, Boltyanskii VG, Gamkrelidze RV, Mishchenko EF (1962) The mathematical theory of optimal processes. Wiley, New York

Raab A (2000) On the Dirac-Frenkel/Mclachlan variational principle. Chem Phys Lett 319(5):674–678

Rabitz H, de Vivie-Riedle R, Motzkus M, Kompa K (2000) Whither the future of controlling quantum phenomena? Science 288:824

Raghunathan S, Nest M (2011) Critical examination of explicitly time-dependent density functional theory for coherent control of dipole switching. J Chem Theory Comput 7(8):2492–2497

Räsänen E, Castro A, Werschnik J, Rubio A, Gross EKU (2007) Optimal control of quantum rings by terahertz laser pulses. Phys Rev Lett 98:157404

Rice SA, Zhao M (2000) Optical control of molecular dynamics. Wiley, New York

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52:997–1000

Shi S, Rabitz H (1989) Selective excitation in harmonic molecular systems by optimally designed fields. Chem Phys 139:185–199

Shi S, Woody A, Rabitz H (1988) Optimal control of selective vibrational excitation in harmonic linear chain molecules. J Chem Phys 88(11):6870

Tannor DJ, Rice SA (1985) Control of selectivity of chemical reaction via control of wave packet evolution. J Chem Phys 83(10):5013–5018

Tannor DJ, Kosloff R, Rice SA (1986) Coherent pulse sequence induced control of selectivity of reactions: exact quantum mechanical calculations. J Chem Phys 85(10):5805–5820

van Leeuwen R (1999) Mapping from densities to potentials in time-dependent density-functional theory. Phys Rev Lett 82(19):3863–3866

van Leeuwen R, Stefanucci G (2013) Nonequilibrium many-body theory of quantum systems. Cambridge University Press, Cambridge

Walkenhorst J, De Giovannini U, Castro A, Rubio A (2016) Tailored pump-probe transient spectroscopy with time-dependent density-functional theory: controlling absorption spectra. Eur Phys J B 89(5):128

Weiner AM (2000) Femtosecond pulse shaping using spatial light modulators. Rev Sci Instrum 71(5):1929–1960

Zhu W, Rabitz H (1998) A rapid monotonically convergent iteration algorithm for quantum optimal control over the expectation value of a positive definite operator. J Chem Phys 109(2):385–391

# Part IV
# Atomistic Simulations

# Atomistic Simulations: An Introduction

# 23

## Wanda Andreoni and Sidney Yip

## Contents

### Abstract

Atomistic simulations are recognized worldwide as an important tool for deepening our understanding of the physics, chemistry, and mechanics of materials. Formidable efforts are being made in the theoretical, modeling, and computational approaches to extend their validity to the complex systems and processes relevant to current technologies as well as to enhance their predictive power. This section reports on recent developments which address the problems of system-size and time scales that are especially critical for ab initio atomistic simulations. In particular, three chapters are devoted to the remarkable advances recently achieved in enhanced sampling methods that are paving the way to the study of the dynamics of complicated transformations and chemical reactions in condensed-matter systems.

W. Andreoni (✉)
Institute of Physics, Swiss Federal Institute of Technology – Lausanne, Lausanne, Switzerland
e-mail: wanda.andreoni@epfl.ch

S. Yip (✉)
Department of Nuclear Science and Engineering, Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: syip@mit.edu

# 1 Introduction

During the past two decades, computer simulations at the atomic level have become the dominant tool for the investigation of physical and chemical processes in real materials. Still, many are the challenges one faces in trying to render the simulation "realistic." For example, building a "reliable" model of the system to be simulated is a task of primary importance. On the one hand, the model size should be adequate to represent its composition and the range of the relevant interactions. On the other hand, an approximation must be selected for the system Hamiltonian. Currently, the description of the interatomic interactions mainly relies either on semi-empirical potentials or on "standard" (GGA) schemes derived from density functional theory (DFT). In several cases, these approximations are not sufficiently accurate and more sophisticated methods are required, which are computationally expensive. These include DFT-based molecular dynamics using hybrid exchange-correlation functionals. For advances regarding calculations of the electronic structure of extended systems we refer the reader to the MTM Section "Electronic Structure of Materials by Ab Initio Methods" (Editor: Rubio).

Moreover, there exist fundamental challenges pertaining to the simulation methods themselves, be they Monte Carlo or molecular dynamics. Their ability to represent the thermodynamics as well as the kinetics of activated processes remains the main concern, and in particular their aptness to identify and characterize the so-called rare events. A general discussion of related issues and open problems can be found in a very recent review (Camilloni and Pietrucci 2018).

In this section important method developments of the last decade, including very recent proposals, are discussed which are expected to play an important role for many applications in the near future. A brief overview is reported here with the aim of guiding the reader through a diverse and highly relevant set of chapters.

# 2 A Brief Chapter Overview

The "system-size problem" is especially crucial for nonempirical simulations that involve the calculation of the electronic structure. ▶ Chapter 24, "Extending the Scale with Real-Space Methods for the Electronic Structure Problem" by Chelikowsky shows that a remarkable extension of the size limits can be obtained using a real-space representation of the electronic states, at least within schemes based on density functional theory and reduced to the valence electron subspace via the pseudopotential scheme.

Going beyond "standard" (LDA or GGA) approximations of density functional theory for the description of the electronic structure is important for the atomistic simulations of diverse types of materials. However, currently, this effort is highly computationally demanding. An ingenious and efficient algorithm allowing ab initio molecular dynamics based on correlated wavefunction-based methods is described

in ▶ Chap. 25, "MP2- and RPA-Based Ab Initio Molecular Dynamics and Monte Carlo Sampling" by Hutter et al. A very interesting application to liquid water is then presented including comparison with DFT-based calculations using hybrid exchange-correlation functionals.

In the search for ways to alleviate the "time-scale problem," methods aimed at directly accelerating the simulation are of special interest. For a broad discussion of the latter – including related technical issues – we refer the reader also to the MTM section on "Computational Methods for Long-Timescale Atomistic Simulations" (Editors: Perez and Uberuaga), and to the Chap. 11, "Long Time-Scale Atomistic Modeling and Simulation of Deformation and Flow in Solids" by Fan and Cao in the ACE section "Plenary Topics" (Editors: Andreoni and Yip).

In the present section, Mazzola and Sorella describe in detail a new, very promising, efficient algorithm (▶ Chap. 26, "Accelerated Molecular Dynamics for Ab Initio Electronic Simulations") which is based on Langevin dynamics. A few examples are presented to illustrate its performance, both within classical and ab initio molecular dynamics. We emphasize that in the latter forces can be derived not only from DFT calculations but also from first-principles approaches like the quantum Monte Carlo method.

Enhanced sampling approaches can be considered also as ways to solve the "time-scale problem," when one is interested to investigate systems with multiple metastable states separated by high-energy barriers and capture the mechanisms leading to the related transitions.

In 2002, the seminal paper by Laio and Parrinello (2002) introduced metadynamics as an efficient tool for configurational sampling and the reconstruction of the free-energy surface. In particular, it was readily implemented in DFT-based molecular dynamics. Several ways to extend and improve on the original scheme have been proposed since then. ▶ Chapter 27, "Metadynamics: A Unified Framework for Accelerating Rare Events and Sampling Thermodynamics and Kinetics" by Bussi, Laio, and Tiwari, provides an overview of this progress. In particular, we would like to draw attention to the discussion of a simple algorithm based on metadynamics that was recently introduced by Tiwari and Parrinello (2013) to determine kinetic rates and later successfully applied in several cases. This approach is a welcome addition to the toolbox of methods addressing the kinetics - something that remains one of the major challenges for molecular simulation.

Two additional chapters report on more recent developments and results.

Pietrucci's ▶ Chap. 28, "Novel Enhanced Sampling Strategies for Transitions Between Ordered and Disordered Structures," introduces an effective representation of the physical system, namely the adjacency matrix of interatomic connections. It goes on to demonstrate the ability of this representation to characterize the different states (phases) of the system and to lead to transitions from one to the other. Selected examples illustrate the results of applications of this new conceptual framework coupled to ab initio molecular dynamics and metadynamics or umbrella-sampling. We emphasize that this scheme not only overcomes the bottleneck of having to

choose ad hoc collective variables, but is applicable to a wide variety of physical and chemical processes. Moreover, it describes ordered and disordered systems on the same footing and thus naturally allows for the simulation of order–disorder phase transitions.

▶ Chapter 29, "Variationally Enhanced Sampling" by Valsson and Parrinello, describes a novel theoretical framework aimed at establishing rigorous foundations for enhanced sampling. The problem of exploring and estimating free energy landscapes is recast into that of minimizing a convex functional. In particular, this new scheme allows one to address systems characterized by high-dimensional free energy surfaces or complex processes that cannot be described with a small number of (collective) variables.

Parallel to the many developments in ab initio simulations, progress continues to be made in devising interatomic potentials for a wide variety of materials. This is important because a fully classical scheme permits simulations using larger-size models and for longer duration. On the other hand, other severe fundamental problems continue to limit their application in spite of the increasing development of specific software. Force fields tailored to simulate biological systems have undoubtedly been very successful. For other materials – of interest to this handbook – however, their accuracy and predictive power is often not satisfactory. Typically, in the case of analytic potentials, the functional form is fixed *a priori* and the parameters are adjusted to either empirical data or some DFT calculations or to a mixed set of results, i.e. from both experiment and computations. A general criticism to the use of classical potentials – including the widely employed class of "reactive potentials" (see, e.g., Liang et al. 2013) – regards their transferability, namely their validity in modeling a system in physical and chemical conditions different from those considered in the fitting procedure.

Two examples based on classical schemes are presented in this section.

One is specifically targeted to water. Indeed, building a classical potential that could reproduce a variety of its (anomalous) properties has been the focus of research for decades (for a recent review see, e.g., Demerdash et al. 2018). Not surprisingly, the physical behavior of water continues to be under constant investigation, in diverse conditions and environments, not only experimentally but also via molecular simulations. ▶ Chapter 30, "Water: Many-Body Potential from First Principles (From the Gas to the Liquid Phase)" by Paesani illustrates the performance of a recently developed potential (MB-pol) by means of a critical comparison with ab initio calculations and experiment.

Significant progress in the creation of classical approaches was marked by the introduction of high-dimensional neural network potentials (NNP) for the prediction of the potential energy surface (Behler and Parrinello 2007). ▶ Chapter 31, "Neural Network Potentials in Materials Modeling" by Hellström and Behler provides an exhaustive and critical overview of the state of the art and clarifies the steps needed in the construction of NNP, including their validation. Very recently, the combination of artificial neural networks and advanced machine learning strategies, also employing accurate reference data, has led to a classical scheme of unprecedented performance for the simulation of liquid and solid water (Cheng et al. 2019).

# 3   Conclusions

Over the last 20 years, atomistic simulations have become a "must" for progress to be made in many domains of materials science. Their ability is called for not only for understanding the behavior of given materials under certain conditions, but also for predicting it as well as for designing novel materials and devising or modifying processes relevant to technology. In order to comply with these "tasks," extension of computational power is essential but not sufficient. Currently, advances are primarily required in the simulation methods, in particular along the lines discussed in the various articles of this section. On the one hand, we repeat that the endeavor to improve on the accuracy of the description of the interatomic interactions is crucial, and also to increase the size-scale of the molecular model and its complexity (composition, structure, etc.) and the timescale of the events to be simulated. On the other hand, we recall some of the problems that are still beyond the reach of atomistic simulations: the characterization of the kinetics of nontrivial chemical reactions, the evolution of fundamental processes like that of the nucleation of, say, a crystal phase from a solution, or the dynamics of interfaces between heterogeneous materials as present in typical devices and their structural transformations under device operating conditions. At this stage, not only extensions of current methodologies but also new formulations appear to be necessary.

# References

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98:146401

Camilloni C, Pietrucci F (2018) Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. Adv Phys X 3:1477531

Cheng B, Engel EA, Behler J, Dellago C, Ceriotti M (2019) Ab initio thermodynamics of liquid and solid water. Proc Natl Acad Sci 116:1110

Demerdash O, Wang L-P, Head-Gordon T (2018) Advanced models for water simulations. WIREs Comput Mol Sci 8:e1355

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci 99:12562

Liang T, Shin YK, Cheng Y-T, Yilmaz DE, Vishnu KG, Verners O, Zou C, Phillpot SR, Sinnott SB, van Duin ACT (2013) Reactive potentials for advanced atomistic simulations. Annu Rev Mater Res 43:109

Tiwary P, Parrinello M (2013) From metadynamics to dynamics. Phys Rev Lett 111:230602

# Extending the Scale with Real-Space Methods for the Electronic Structure Problem

# 24

James R. Chelikowsky

## Contents

**Abstract**

In principle, the electronic structure of a material can be determined by a solution of the many-body Schrödinger equation. This was first noted by Dirac shortly after the invention of quantum mechanics in 1929. However, Dirac also noted that the solution of the many-body quantum mechanical equations was much too difficult to be solved. He challenged his colleagues to develop "practical methods of applying quantum mechanics," which can lead to an explanation of the main features of complex atomic systems." In this chapter, we explore concepts and algorithms targeting "Dirac's challenge." Two key physical concepts will be employed: pseudopotential theory and density functional theory. For many weakly correlated systems, this formalism works well for ground-state properties such as phase stability, structural properties, and vibrational modes. However,

J. R. Chelikowsky (✉)
Center for Computational Materials, Institute for Computational Engineering and Sciences, Departments of Physics and Chemical Engineering, The University of Texas at Austin, Austin, TX, USA
e-mail: jrc@utexas.edu

applying this approach to large systems, e.g., systems with thousands of atoms, remains a challenge even with contemporary computational platforms. The goal of this chapter is to show how new algorithms can be used to extend computations to these systems. The approach centers on solving the nonlinear Kohn-Sham equation by a nonlinear form of the subspace iteration technique. This approach results in a significant speedup, often by more than an order of magnitude with no loss of accuracy. Numerical results are presented for nanoscale systems with tens of thousands of atoms and new methods are proposed to extend our work to even larger systems.

## 1    Introduction

Many properties of a material can be determined from a knowledge of its electronic structure, i.e., the energy and spatial distribution of electronic states. For example, optical and structural properties can be predicted by knowing the energetic and spatial distribution of electrons in the material.

This was recognized with the inception of quantum mechanics shortly after the work of Schrödinger, Heisenberg, and Dirac. As Dirac noted: "The underlying physical laws [quantum theory] necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed" (Dirac 1929).

The most popular electronic structure methods are based on pseudopotentials (Phillips 1958; Phillips and Kleinman 1959; Chelikowsky and Cohen 1992) and density functional theory (Hohenberg and Kohn 1964; Kohn and Sham 1965). Pseudopotential theory simplifies the electronic structure problem by replacing the "all electron" atomic potential with an effective "pseudopotential," which only binds the valence states. Density functional reduces the original multi-electron Schrödinger equation into an effective one-electron Kohn-Sham equation, where the nonclassical electronic interactions are replaced by a functional of the charge density, i.e., the spatial distribution of electrons. Combining pseudopotentials with DFT can significantly reduce the number of computed one-electron wave functions. More importantly the energy and length scales are set solely by the valence states. Species such as a silicon and tin can be treated on equal footing in solving the electronic structure problem even though the number of electrons in tin is more than three times that of silicon.

Even with these practical simplifications, solving the Kohn-Sham equation remains computationally challenging when the system of interest contains a large number of atoms, e.g., more than a few thousand atoms. Researchers have advocated different approaches to overcome the computational impediments (Saad et al. 2010). These approaches may be classified as basis-free or basis-dependent approaches, according to whether they use an explicit basis set for electronic orbitals or not.

Among the most popular basis-dependent approaches is the plane wave basis set, which is frequently used in applications of pseudopotential density functional theory to periodic systems where plane waves can easily accommodate the boundary conditions (Kresse and Furthmülle 1996; Giannozzi et al. 2009). Another popular approach uses localized basis sets such as Gaussian orbitals, which are commonly utilized in quantum chemistry calculations (Koch and Holthausen 2000; Martin 2004; Saad et al. 2010).

We focus on a different approach based on *real-space* methods, which are "basis free." These methods are popular owing in great part to their great simplicity and ease of implementation (Chelikowsky et al. 1994a, b; Seitsonen et al. 1995; Beck 2000; Fattebert and Bernholc 2000; Ono and Hirose 2005; Kronik et al. 2006; Enkovaara et al. 2010; Andrade et al. 2013). They can be readily implemented in parallel computing environments as they minimize global communications (Saad et al. 2010). Another advantage of real-space methods is that they do not impose artificial periodicity to handle nonperiodic systems. While plane wave basis techniques can be applied to clusters (or molecules), such applications often proceed by placing the cluster of interest in a large supercell. Provided the supercell is sufficiently large so that the cluster is removed from neighboring replicants, the electronic structure solution will correspond to that of the isolated cluster. However, because the potential from neighboring cells can be significant, in some cases this is not a practical solution, i.e., supercell solutions may converge slowly with the size of the cell (Alemany et al. 2004). A related, and perhaps more significant, issue is that the use of supercells is not easily applied to systems that are electronically charged. Charged systems can be handled within plane wave methods by including a compensating uniform charge or a more complex procedure (Alemany et al. 2004; Teter et al. 1992). Real-space methods need not address such complications. The application of the Hamiltonian to electron wave functions is performed entirely in real space. While the Hamiltonian matrix in real-space methods is typically larger than similar applications with plane waves, the matrices are *extremely sparse* and never stored or computed explicitly. Only matrix-vector products that represent the application of the Hamiltonians on wave functions are calculated.

The prime obstacle to solving the Kohn-Sham equations is obtaining a solution to the nonlinear eigenvalue problem. While one can judiciously choose a well-crafted basis or pseudopotential, the end game is a solution of a difficult eigenvalue problem. For example, suppose one considered a local basis of ten orbitals per atom for a system containing 10,000 atoms. The resulting Hamiltonian matrix is large ($100,000 \times 100,000$) and dense. While we might be interested in obtaining a small fraction of the total number of eigenvalue solutions, this is a difficult problem. Moreover, the scaling of the problem is difficult with a solution often exceeding $\mathcal{O}(N^n)$ where $N$ is the number of eigenvalues and $n$ is formally equal to 3, but owing to a small prefactor in realistic computations the operational value is closer to 2 (Saad et al. 2010). Often workers will focus solely on the scaling, but the prefactor is important too. The prefactor for orthogonalization is often small, so the resulting operation is closer to the square of the eigenvalue number. In general, the key measure is not scalability per se but *time to solution* [TTS].

In this chapter, we center on efficient algorithms for eigensolvers applicable to real-space Hamiltonians. Specifically, we will present examples of nonlinear Chebyshev-filtered subspace iteration algorithms implemented in our solution package called PARSEC (Pseudopotential Algorithm for Real-Space Electronic Calculations) (Chelikowsky et al. 1994a, b).

Although described in the framework of real-space pseudopotentials, subspace filtering can be employed to other approaches. This method takes advantage of the fact that intermediate self-consistent iterations do not require accurate eigenvalues and eigenvectors of the Kohn-Sham equation. Still, the self-consistent solution is the *same* accuracy as other eigensolvers. Unlike some so-called "$\mathcal{O}(N)$" methods (Otsuka et al. 2008; Goedecker 1999), subspace filtering is equally applicable to metals and insulators.

Solutions by filtering do not focus on the intermediate linearized Kohn-Sham eigenvalue problems. Typically, explicit eigenvectors are computed only at the first self-consistent field (SCF) iteration, in order to provide a viable initial subspace. After the first self-consistent step, the explicit computation of eigenvectors at each succeeding iteration is replaced by a subspace filtering step. The method reaches self-consistency within a similar number of iterations, when compared to eigenvalue-based approaches. However, since eigenvalues are not explicitly computed after the first step, a significant gain in execution time results when compared with methods based on explicit diagonalization. When compared with calculations based on efficient eigenvalue packages such as ARPACK (Lehoucq et al. 1998) and TRLan (Wu et al. 1999; Wu and Simon 2000), an order of magnitude speedup is often observed.

Filtering methods enable us to perform a class of challenging electronic structure calculations, including clusters with over tens of thousands atoms, which were not feasible before without invoking additional approximations in the Kohn-Sham problem (Zhou et al. 2006a, b; Zhou and Saad 2007; Schofield et al. 2012b).

## 2    The Kohn-Sham Equation

Density functional theory provides a practical method for solving the electronic structure problem, the Kohn-Sham equation (Kohn and Sham 1965):

$$\left[ -\frac{\hbar^2}{2m}\nabla^2 + V_{total}(\rho(r), r) \right] \Psi_i(r) = E_i \Psi_i(r) \tag{1}$$

where $\Psi_i(r)$ is a wave function or eigenfunction, $E_i$ is a Kohn-Sham eigenvalue, $\hbar$ is the Planck constant, and $m$ is the electron mass. (We use atomic units: $\hbar = m = e = 1$ in the following discussion.)

The *total potential*, $V_{total}$, is the sum of three contributions:

$$V_{total}(\rho(r), r) = V_{ion}(r) + V_H(\rho(r), r) + V_{xc}(\rho(r), r), \qquad (2)$$

where $V_{ion}$ is the external ionic pseudopotential, $V_H$ is the Hartree or Coulomb potential, and $V_{xc}$ is the effective exchange-correlation potential. Within density functional theory, exchange-correlation potentials depend solely on the *charge density* $\rho(r)$, which is defined as

$$\rho(r) = \sum_{i=1}^{n_{occ}} |\Psi_i(r)|^2. \qquad (3)$$

Here $n_{occ}$ is the number of occupied Kohn-Sham levels, which is equal to half the number of valence electrons in the system if the system is nonmagnetic. Equation (3) can be generalized to situations where the highest occupied states have fractional occupancy or when there is an imbalance in the number of electrons for each spin component. The Hartree potential is solved from Poisson's equation using the electronic charge density.

The total potential, $V_{total}$, depends on the charge density $\rho(r)$, which in turn depends on the wave functions $\Psi_i$. The Kohn-Sham equation can be viewed as a *nonlinear eigenvalue problem*.

Such equations are typically solved using an iterative procedure to generate a "self-consistent field." The iteration process begins with an initial guess of the charge density, which is usually constructed from a superposition of free atomic charge densities. Using this approximate density, one obtains the initial $V_{total}$ and solves Eq. (1) for $\Psi_i(r)$'s to update $\rho(r)$ and $V_{total}$. Then the Kohn-Sham (Eq. (1)) is solved again for the new $\Psi_i(r)$'s, and the process is iterated. We terminate the process when the charge density is self-consistent, i.e., the solution of the Kohn-Sham equation yields a charge density consistent with the Hamiltonian used to generate the charge density.

The standard SCF process is described in Algorithm 1 and illustrated in Fig. 1

---

**Algorithm 1** Self-consistent-field iteration

1. Provide initial guess for $\rho(r)$ and get $V_{total}(\rho(r), r)$.
2. Solve for $\Psi_i(r)$, $i = 1, 2, \ldots$, from

$$\left[ -\frac{1}{2} \nabla^2 + V_{total}(\rho(r), r) \right] \Psi_i(r) = E_i \Psi_i(r).$$

3. Compute the new charge density $\rho(r) = \sum_{i=1}^{n_{occ}} |\Psi_i(r)|^2$.
4. Obtain new Hartree potential $V_H$ by solving: $\nabla^2 V_H(r) = -4\pi\rho(r)$.
5. Update $V_{xc}$; get new $\tilde{V}_{total}(\rho, r) = V_{ion}(r) + V_H(\rho, r) + V_{xc}(\rho, r)$ with a potential-mixing step.
6. If $\|\tilde{V}_{total} - V_{total}\| < tol$, stop; Else, $V_{total} \leftarrow \tilde{V}_{total}$, go to step 2.

---
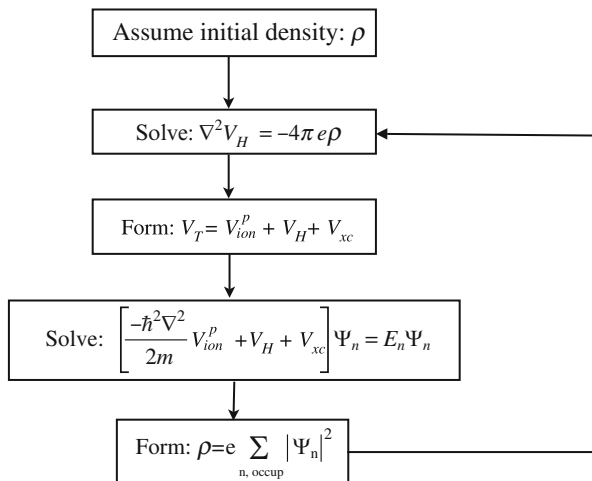
**Fig. 1** Flow diagram for obtaining a self-consistent solution of the Kohn-Sham equation

The number of eigenvectors needed in *Step 2* of Algorithm 1 is just the number of occupied states. In practice, a few more eigenvectors are usually computed. For complex systems, i.e., when the number of valence electrons is large, each of the linearized eigenvalue problems can be computationally demanding. This is compounded by the fact that Hamiltonian matrices can be of very large size. For example, a cubic system with 100 grid points on a size results in a matrix of $10^6 \times 10^6$, although the matrix remains very sparse.

For this reason, we wish to reduce the computational load in *Step 2* of Algorithm 1. There are several avenues open to us. One could use some physical arguments to reduce the matrix size or zero some existing elements. One could attempt to avoid diagonalization altogether, as is done in work represented by linear-scaling or order-N methods (see, e.g., Otsuka et al. 2008; Goedecker 1999). This approach, however, has other limitations. In particular, the approximations involved rely heavily on some decay properties of the density matrix in certain function bases. In particular, they can be difficult to implement in real-space discretizations or for systems where the decay properties are not optimal, e.g., in metals. Another option is to use better (faster) eigensolvers.

The approach discussed here avoids *standard* diagonalizations but otherwise makes no new approximations to the Hamiltonian. We take advantage of the fact that accurate eigenvectors are unnecessary at each SCF iteration, since Hamiltonians are only approximate in the intermediate SCF steps, and exploit the nonlinear nature of the problem. The main point of the new algorithm is that once we have a good starting point for the Hamiltonian, it suffices to *filter* each basis vector at each iteration. In the intermediate SCF steps, these vectors are no longer eigenvectors, but together they represent a good basis of the desired invariant subspace.

## 3      Real-Space Approach

A real-space solution of the Kohn-Sham equation is commonly approached using a uniform, Cartesian grid as shown in Fig. 2.

We solve the Kohn-Sham equation, Eq. 1, for the grid-sampled values of wave functions, potentials, and the electron density. The Kohn-Sham wave functions are vectors whose length is the number of grid points. Each of the operators in the Kohn-Sham equation (kinetic, pseudopotential, Hartree, and exchange-correlation) is represented by a square matrix.

We use high-order finite differencing to express the kinetic energy Laplacian term. The discretization of the Laplacian utilizes *several/*neighbors for calculating the derivative around each point of interest, instead of just the immediate neighbors, i.e.:

$$\nabla^2 \varphi(x, y, z)$$

$$= \frac{1}{h^2} \sum_{l,k,m=-M}^{M} [C_k \varphi(x+kh, y, z) + C_l \varphi(x, y+lh, z) + C_m \varphi(x, y, z+mh)]$$

(4)

where $M$ is half the number of neighbors used along each axis or half the *expansion order* and $h$ is the grid spacing. The coefficients $(C_k, C_l, C_m)$ depend only on the choice of the expansion order, $M$. They can be found from a high-order Taylor expansion, although other algorithms for their determination are also available (Fornberg and Sloan 1994). The kinetic energy matrix has only $3M$ elements in each row (in addition to the diagonal), resulting in a very sparse matrix.

**Fig. 2** The finite-difference real-space approach. For open boundary conditions (shown in the figure), wave functions are sampled on a uniform grid within a spherical domain and vanish outside its boundary. Black dots denote grid points, and the blue balls denote atoms (a dimer example is shown)
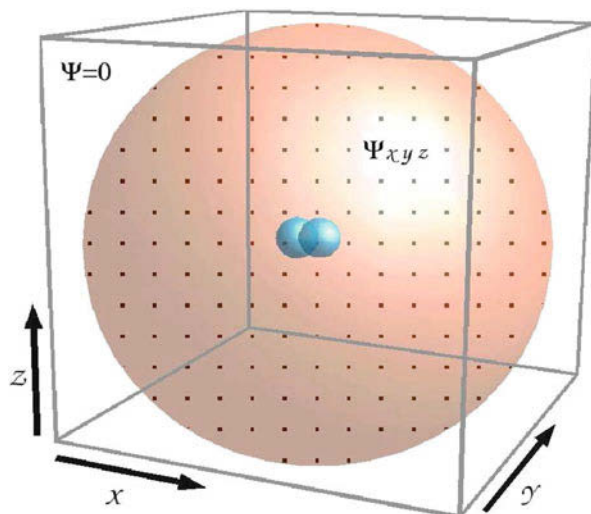
There are many algorithms for constructing reliable pseudopotentials (Phillips 1958; Phillips and Kleinman 1959; Chelikowsky and Cohen 1992). We employ norm-conserving pseudopotentials. Typically Troullier-Martins pseudopotentials (Troullier and Martins 1991) are used. These potentials have been optimized for convergence with a plane wave basis and can be easily implemented in real-space methods.

The price paid for the simplification afforded by pseudopotentials is that the ionic potential operator is no longer local. Different atomic orbitals experience different pseudopotentials, so that application of the pseudopotential requires nonlocal projectors (Chelikowsky and Cohen 1992):

$$V_{ps}(\mathbf{r}) = \sum_{a=1}^{N} \sum_{l} V_{a,ps}^{l}(|\mathbf{r} - \mathbf{R}_a|) \sum_{m=-l}^{l} |l, m \rangle\langle l, m| \tag{5}$$

where $V_{a,ps}^{l}$ is the pseudopotential of the $l$th angular momentum orbital of the $a$th atom, $\mathbf{R}_a$ is the coordinate of the $a$th atom, and $|l, m\rangle$ is the $lm$th spherical harmonic. An easy way of making the matrix corresponding to the pseudopotential operator sparse is to use the projection scheme suggested by Kleinman and Bylander (1982):

$$V_{ps(\mathbf{r})} = \sum_{a=1}^{N} \left[ V_{a,ps}^{loc}(\mathbf{r}) + \frac{|\Delta V_{a,ps}^{l}(\mathbf{r})\varphi_{lm}^{a}(\mathbf{r})\rangle\langle\Delta V_{a,ps}^{l}(\mathbf{r})\varphi_{lm}^{a}(\mathbf{r})|}{\langle\varphi_{lm}^{a}(\mathbf{r})|\Delta V_{a,ps}^{l}(\mathbf{r})|\varphi_{lm}^{a}(\mathbf{r})\rangle} \right], \tag{6}$$

where $V_{a,ps}^{loc}(\mathbf{r})$ is a pseudopotential corresponding to one specific angular momentum component (of atom a), arbitrarily chosen as the local one, and $\Delta V_{a,ps}^{l}(\mathbf{r}) \equiv V_{a,ps}^{l}(\mathbf{r}) - V_{a,ps}^{loc}(\mathbf{r})$. Finally, $\varphi_{lm}^{a}(\mathbf{r})$ is the atomic pseudo-wave function with $lm$ quantum angular momentum numbers.

This form of the pseudopotential has several advantages: $\Delta V_{a,ps}^{l}$ differs from zero only inside a relatively small region around each atom. (The size of the region corresponds to the pseudopotential cutoff radius). Only the normalization scalars $c_{lm}^{a} = \langle\varphi_{lm}^{a}(\mathbf{r})|\Delta V_{a,ps}^{l}(\mathbf{r})|\varphi_{lm}^{a}(\mathbf{r})\rangle$ and the highly sparse vectors $\mathbf{u}_{lm}^{a} = |\Delta V_{a,ps}^{l}(\mathbf{r})\varphi_{lm}^{a}(\mathbf{r})\rangle$ are stored explicitly instead of holding an entire matrix. The $c_{lm}^{a}$ and $\mathbf{u}_{lm}^{a}$ terms are computed only once in the entire self-consistent solution, because they only depend on the atom coordinates, $\mathbf{R}_a$, and angular momentum numbers, $\{l, m\}$, which do not change during the iterative computation.

The Hartree potential is found by solving the Poisson equation:

$$\nabla^2 V_H(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{7}$$

using the conjugate gradient method. The Hartree potential is local with the associated matrix diagonal. The exchange-correlation potential is defined as the functional derivative of the exchange-correlation energy with respect to the charge density. For a local exchange-correlation, the associated matrix is diagonal.

**Fig. 3** Schematic of the Hamiltonian matrix for $Si_{28}H_{36}$ showing the non-zero matrix elements for the kinetic and nonlocal pseudopotential contributions
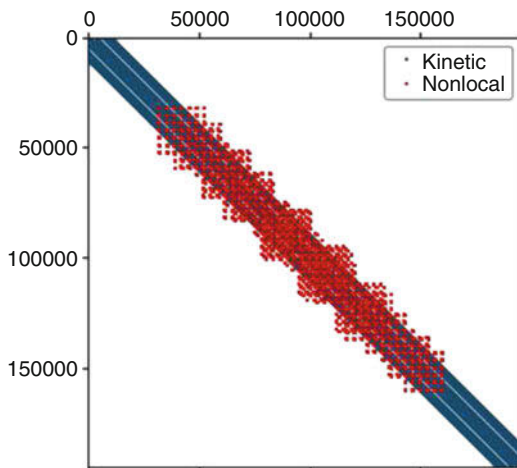
The resulting Hamiltonian matrix, being a sum of all matrix operators, each of which is diagonal or highly sparse, is sparse itself. The only quantities computed and stored explicitly are the diagonal elements (diagonal component of kinetic energy, local pseudopotential, Hartree potential, and exchange-correlation potential), the high-order finite-difference expansion coefficients, and the relatively small nonlocal pseudopotential vectors $\mathbf{u}_{lm}$. This results in huge memory savings, which are crucial for the computation of large systems containing thousands of atoms or more.

The structure of the Hamiltonian matrix is illustrated in Fig. 3. Non-zero matrix elements for the kinetic energy operator and for the nonlocal pseudopotential elements are shown for a small nanocrystal ($Si_{28}H_{36}$). The matrix size is $195,112 \times 195,112$ and with $0.0118\%$ of the matrix not sparse. The matrix need not be explicitly stored, nor are there off-diagonal element updates.

## 4    Eigensolvers Using Subspace Filtering

Subspace filtering is an effective operation to extract a subset of eigenvalues. We illustrate some of the essential features of the filtering process. Consider a matrix problem: eigenvalues. Consider a matrix problem:

$$H\Psi = \Lambda\Psi \tag{8}$$

where $H$ is the real-space matrix as defined above. $\Psi$ is a wave function matrix for the eigenfunction solutions for an $N \times N$ matrix:

$$\Psi = \begin{pmatrix} \psi_{1,1} & \psi_{1,2} & \cdots & \psi_{1,N} \\ \psi_{2,1} & \psi_{2,2} & \cdots & \psi_{2,N} \\ \vdots & \vdots & \cdots & \vdots \\ \psi_{N,1} & \psi_{N,2} & \cdots & \psi_{N,N} \end{pmatrix} \tag{9}$$

The eigenvalue matrix is given by

$$
\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix} \tag{10}
$$

where the Kohn-Sham eigenvalues are given by $\lambda_i$.

Using this notation, we consider the product $\Phi = \Psi \Psi^T$. The diagonal element of $\Phi$ will give the sum of the squares of the wave functions for each grid point. Now consider a different matrix $\Psi U$ where $U$ is an orthonormal matrix such that $UU^T = I$, where $I$ is the identity matrix. We can write $\Phi = \Psi(UU^T)\Psi^T = (\Psi U)(U^T \Psi^T)$, and using the matrix identity, $B^T A^T = (AB)^T$, we get $\Phi = (\Psi U)(\Psi U)^T$.

This matrix manipulation has a clear implication. We can use either $\Psi$ or $\Psi U$ to find $\Phi$. All we care about is the diagonal of $\Phi$. So, we do not need to know each eigenvector to get the charge density as rotation of the space will do.

We can be more explicit. Suppose we consider to define a few more matrices such as $P(H)$ where $P$ is a polynomial. As a simple example, suppose $P = H^2$. Applying $H^2$ to $\Psi$, we find $H^2 \Psi = H(H\Psi) = H\Lambda \Psi = \Lambda^2 \Psi$. In general we can write $P(H)\Psi = P(\Lambda)\Psi$.

Consider the following operations. Suppose we take an approximate eigenfunction, $\tilde{\psi}_j$, for the $j$ eigenvector and apply a polynomial $P(H)$ to $\tilde{\psi}_j$:

$$
P(H)\tilde{\psi}_j = P(H)\left( \sum_i \alpha_{ij} \psi_j \right) = \sum_i \alpha_{ij} P(\lambda_j)\psi_j \tag{11}
$$

where we have assumed that our approximate eigensolution can be expressed as

$$
\tilde{\psi}_j = \sum_i \alpha_{ij} \psi_j \tag{12}
$$

This follows from the completeness of the solution set. Suppose we take our filtered eigenfunction, $\psi_j^f$, as

$$
\psi_j^f = P(H)\tilde{\psi}_j = \sum_i \alpha_{ij} P(\lambda_j)\psi_j \tag{13}
$$

and chose an appropriate polynomial, e.g., suppose we chose $P(\lambda) \approx 0$ for all values of $\lambda$ above the fermi level. In this case, only components of the occupied wave function or eigenfunctions would contribute. This is what we want.

**Fig. 4** Schematic of a filtering polynomial showing regions of "wanted" and "unwanted" eigenvalues
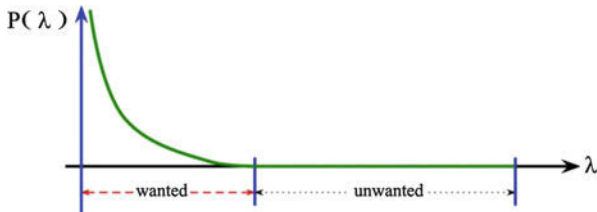


A schematic of a filtering polynomial is shown in Fig. 4. Any polynomial with this shape will do. We want a large amplitude in the desired subspace and a small one otherwise.

## 5 Chebyshev Filters

Often Chebyshev polynomials are useful for our purposes as they have a number of desirable properties. Chebyshev polynomials of the first kind (Parlett 1998; Saad 1992) are defined, for $k = 0, 1, \cdots ,$ :

$$C_k(t) = \begin{cases} \cos\big(k \ \cos^{-1}(t)\big), & -1 \le t \le 1, \\ \cosh\big(k \ \cosh^{-1}(t)\big), & |t| > 1. \end{cases}$$

We can express the polynomials in s convenient recursion relation. We start with the first two polynomials: $C_0(t) = 1$, $C_1(t) = t$. The following recurrence is easy to derive from properties of the cosine function and can be written:

$$C_{k+1}(t) = 2t \ C_k(t) - C_{k-1}(t), \quad t \in \mathbb{R} \tag{14}$$

Within the interval $|t| \le 1$, the polynomials have nodes and possess a magnitude less than unity. In principle, we want the filtering process to magnify the desired components of the wave function. If the process is repeated indefinitely, the filtered wave function should have zero components in the "unwanted" part of the eigenvalue spectrum.

A filtering operation applied to a vector has the effect of enhancing desired eigen-components of the vector relative to undesirable components. If the process is repeated indefinitely, the resulting vector will have zero components in the unwanted part of the spectrum. Suppose our wanted eigenvalues correspond to the occupied states. Our goal is to filter out all components associated with the empty states or, equivalently, to enhance the components associated with occupied states, relative to other components.

We exploit well-known properties of Chebyshev polynomials to accomplish filtering. It is known that among all polynomials of degree $k$, which have value one at a certain point $|\gamma| > 1$, the polynomial $C_k(t)/C_k(\gamma)$ is the one whose maximum absolute value in the interval $[-1, 1]$ is minimal. Thus, $C_k(t)/C_k(\gamma)$ can be viewed

**Fig. 5** Degree six Chebyshev polynomial on the interval [0, 2]. The shaded area corresponds to eigen-components that will be amplified relative to the other eigen-components. The value of $l$ is determined by an affine mapping. See text
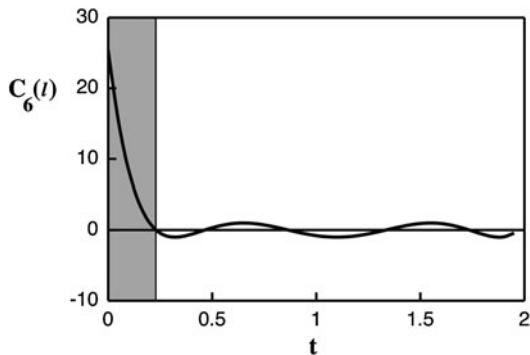


as an optimal polynomial if one wishes to dampen values of the polynomial in $[-1, 1]$ among all polynomials $p$ of degree $k$, scaled so that $p(\gamma) = 1$. A 6th degree Chebyshev polynomial so scaled is shown in Fig. 5.

Assume that the full spectrum of $H$ (denoted by $\Lambda(H)$) is contained in $[\gamma, b]$. Then in order to approximate the eigensubspace associated with the lower end of the spectrum, say $[\gamma, a]$ with $\gamma < a < b$, it is necessary to map $[a, b]$ into $[-1, 1]$ before applying the Chebyshev polynomial. This can be easily realized by an affine mapping.

Assume that the full spectrum of $H$ (denoted by $\Lambda(H)$) is contained in $[\gamma, b]$. In order to approximate the eigensubspace associated with the lower end of the spectrum, say $[\gamma, a]$ with $\gamma < a < b$, it is necessary to map $[a, b]$ into $[-1, 1]$ before applying the Chebyshev polynomial. This can be easily realized by an affine mapping defined as

$$l(t) := \frac{t - c}{e}; \quad c = \frac{a + b}{2}, \quad e = \frac{b - a}{2}$$

where $c$ denotes the center and $e$ the half-width of the interval $[a, b]$.

The Chebyshev iteration utilizing the recurrence in Eq. 14 to dampen values on the interval $[a, b]$ is given in Algorithm 2 (see Zhou et al. 2006b). We write the filtered vector as $Y$, with the unfiltered vector as $X$, and perform the operation:

$$Y = p_m(H)X \tag{15}$$

where $p_m(t) = C_m[l(t)]$ is the mapped Chebyshev polynomial.

The recursion operation is given by

$$X_{j+1} = \frac{2}{e}(H - cI)X_j - X_{j-1}, \quad j = 1, 2, \ldots, m - 1.$$

with $X_0$ given and $X_1 = (H - cI)X_0$.

---

**Algorithm 2** $[Y] = \texttt{Chebyshev\_filter}(X, m, a, b, \gamma)$

Filter column vectors of $X$ by an $m$ degree Chebyshev polynomial in $H$ that dampens on the interval $[a, b]$. Output the filtered vectors in $Y$.

1. $e = (b - a)/2; \quad c = (b + a)/2;$
2. $\sigma = e/(\gamma - c); \quad\quad \sigma_1 = \sigma; \quad\quad \gamma = 2/\sigma_1.$
3. $Y = \frac{\sigma_1}{e}(HX - cX);$
4. For $i = 2 : m$
5. $\quad \sigma_2 = 1/(\gamma - \sigma);$
6. $\quad Y_{\text{new}} = \frac{2\sigma_2}{e}(HY - cY) - \sigma\sigma_2 X;$
7. $\quad X = Y;$
8. $\quad Y = Y_{\text{new}};$
9. $\quad \sigma = \sigma_2;$
10. End For

---

The above iteration is without any scaling. In the case of the interval $[-1, 1]$, we scaled the polynomial by $C_k(\gamma)$ in order to ensure that the value of the polynomial at $\gamma$ equals one. For general intervals, this leads to the scaled sequence of polynomials (Saad 1992)

$$\tilde{X}_j = \frac{C_j[\frac{2}{e}(H - cI)]}{C_j[\frac{2}{e}(\gamma - cI)]} X_0.$$

The scaling factor is $\rho_j = C_j[\frac{2}{e}(\gamma - cI)]$.

This algorithm requires an estimate for $\gamma$ which, in our case, is the smallest eigenvalue of the Hamiltonian. However, since this is used for scaling, for the purpose of avoiding overflow, only a rough value is needed.

The eigen-components associated with eigenvalues in $[a, b]$ will be transformed to small values, while those to the left of $[a, b]$ will be around unity owing to the properties of the Chebyshev polynomials. This is the desired filtering property when computing an approximation to the eigensubspace associated with the lower end of $\Lambda(H)$. As seen in Algorithm 2, a desired filter can be easily controlled by adjusting two endpoints that bound the higher portion of $\Lambda(H)$.

The lower bound can be any value that is larger than the Fermi level but smaller than the upper bound. It can also be a value slightly smaller than the Fermi level; thanks to the monotonicity of the shifted and scaled Chebyshev polynomial on the spectrum of $H$ and the fact that we compute $s > n_{\text{occ}}$ number of Ritz values, the desired lowered end of the spectrum will still be magnified properly with this choice of lower bound.

Since the previous SCF iteration performs a Rayleigh-Ritz refinement step, this step provides a natural approximation for the lower bound $a$. Indeed, we can

---

**Algorithm 3** CheFSI for SCF calculation

1. Start from an initial guess of $\rho(r)$, get $V_{total}(\rho(r), r)$.
2. Solve $\left[-\frac{1}{2}\nabla^2 + V_{total}(\rho(r), r)\right]\Psi_i(r) = E_i\Psi_i(r)$ for $\Psi_i(r)$, $i = 1, 2, \ldots, s$.
3. Compute new charge density $\rho(r) = 2\sum_{i=1}^{n_{occ}}|\Psi_i(r)|^2$.
4. Solve for new Hartree potential $V_H$ from $\nabla^2 V_H(r) = -4\pi\rho(r)$.
5. Update $V_{xc}$; get new $\tilde{V}_{total}(\rho, r) = V_{ion}(r) + V_H(\rho, r) + V_{xc}(\rho, r)$ with a potential-mixing step.
6. If $\|\tilde{V}_{total} - V_{total}\| < tol$, stop; Else, $V_{total} \leftarrow \tilde{V}_{total}$ (update $H$ implicitly),
   call the Chebyshev-filtered subspace method (Algorithm 4) to get $s$ approximate wave functions; go to step 3.

---

**Algorithm 4** Chebyshev-filtered Subspace (CheFS) method

1. Get the lower bounds $b_{low}$ and $\gamma$ from previous Ritz values (use the largest one and the smallest one, respectively).
2. Compute the upper bound $b_{up}$ of the spectrum of the current discretized Hamiltonian $H$.
3. Perform Chebyshev filtering (call Algorithm 2 in Sect. 5) on the previous basis $\Psi$, where $\Psi$ contains the discretized wave functions of $\Psi_i(r)$, $i = 1, \ldots, s$:
   $\Psi = \texttt{Chebyshev\_filter}(\Psi, m, b_{low}, b_{up}, \gamma)$.
4. Ortho-normalize the basis $\Psi$ by iterated Gram-Schmidt.
5. Perform the Rayleigh-Ritz step:
   (a) Compute $\hat{H} = \Psi^T H \Psi$;
   (b) Compute the eigendecomposition of $\hat{H}$: $\hat{H}U = UD$,
       where $D$ contains non-increasingly ordered eigenvalues of $\hat{H}$ and $U$ contains the corresponding eigenvectors;
   (c) "Rotate" the basis as $\Psi := \Psi U$; return $\Phi$ and $U$.

---

simply take the largest Rayleigh quotient from the previous SCF iteration step as an approximation to the lower bound for the current Hamiltonian.

The upper bound for the spectrum (denoted by $b$) can be estimated by a $k$-step standard Lanczos method. The higher endpoint $b$ must be a bound for the full spectrum of $H$. This is because the Chebyshev polynomial also grows fast to the right of $[-1, 1]$ (Zhou and Saad 2007). So if $[a, b]$ with $b < \lambda_{max}(H)$ is mapped into $[-1, 1]$, then the $[b, \lambda_{max}(H)]$ portion of the spectrum will also be magnified, which will cause the procedure to fail. Therefore, it is imperative that the bound $b$ be larger than $\lambda_{max}(H)$.

## 6    Subspace Iteration Using Filtering

The main structure of Chebyshev filtering, which is given in Algorithm 3, is quite similar to that of the standard SCF iteration (Algorithm 1). One major difference is that the inner iteration for diagonalization at *Step 2* is now performed only at the first SCF step. Thereafter, diagonalization is replaced by a single Chebyshev subspace filtering step, performed by calling Algorithm 4. The filtering process is illustrated in Fig. 6.

Although the charge density (Eq. (3)) requires only the lowest $n_{occ}$ states, the number of computed states, which is the integer $s$ in Algorithm 3, is typically set to
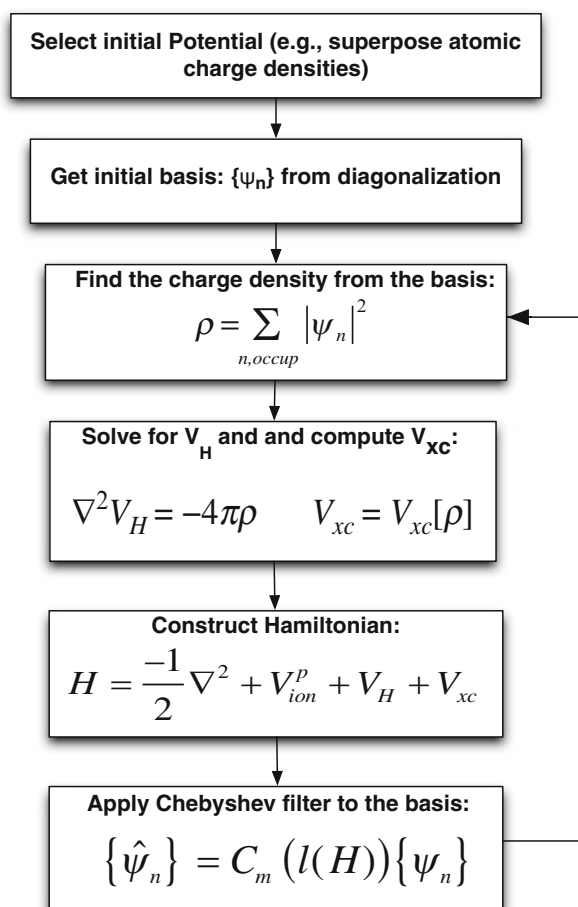


**Fig. 6** Flow diagram for obtaining a self-consistent solution of the Kohn-Shan equation using damped Chebyshev subspace filtering

a value larger than $n_{occ}$, in order to avoid missing any occupied states. In practice we fix an integer $n_{state}$ which is slightly larger than $n_{occ}$.

The parallel implementations of Algorithms 3 and 4 are quite straightforward. The matrix-vector products related to filtering, computing upper bounds, and Rayleigh-Ritz refinement can easily execute in parallel. The re-orthogonalization at *Step 4* of Algorithm 4 uses a parallel version of the iterated Gram-Schmidt DGKS method (Daniel et al. 1976), which scales better than the standard modified Gram-Schmidt algorithm.

If a standard iterative diagonalization method is used to solve the linearized eigenproblem (Eq. 2) at each SCF step, then it also requires (i) the orthonormalization of a (typically larger) basis, (ii) the eigendecomposition of the projected Rayleigh quotient matrix, and (iii) the basis refinement (rotation). These operations need to be performed several times within this single diagonalization. But Algorithm 4 performs each of these operations only once per SCF step. Therefore, although Algorithm 4 scales in a similar way to standard diagonalization-based methods, the scaling constant is much smaller. For large problems, CheFS can achieve a tenfold or more speedup per SCF step, over using the well-known efficient eigenvalue packages such as ARPACK (Lehoucq et al. 1998) and TRLan (Wu et al. 1999; Wu and Simon 2000).

In summary, a standard SCF method has an outer SCF loop – the usual nonlinear SCF loop – and an inner diagonalization loop, which iterates until eigenvectors are within specified accuracy. Algorithm 3 essentially bypasses the second loop, or rather it merges it into a single outer loop, which can be considered as a *nonlinear subspace iteration algorithm*. The inner diagonalization loop is replaced by a single Chebyshev subspace filtering step.

## 7    Applications and Performance

Our real-space package (PARSEC (http://parsec.ices.utexas.edu/)) has been applied to study a wide range of material systems, many of which would not be accessible without a filtering algorithm (Dalpian and Chelikowsky 2006; Tiago et al. 2006; Alemany et al. 2007; Chan et al. 2008, 2009; Chan and Chelikowsky 2010; Chan et al. 2014; Chelikowsky et al. 2011; Sakai et al. 2016).

We focus on large hydrogenated Si nanocrystals, up to systems with over 20,000 atoms. Such a Si nanocrystal is illustrated in Fig. 7. The surface is capped with hydrogen atoms to saturate the surface atoms and remove any dangling bond states from the gap.

Relatively few numerical results exist because of the difficulty of eigenvector-based methods applied to large systems without periodicity. In particular, Zhao et al. (2004) have examined clusters containing up to 1,100 silicon atoms, using the well-known VASP code (Kresse and Furthmülle 1996). They found applications to clusters with more than $\sim$1,200 silicon atoms were "too computationally intensive" with the computation platforms available to them. As a comparison, PARSEC using the Chebyshev filter algorithm (CheFSI), together with the currently developed

**Fig. 7** $Si_{10701}H_{1996}$ nanocrystal the surface is capped with hydrogen atoms shown as a black dots



$Si_{10701}H_{1996}$
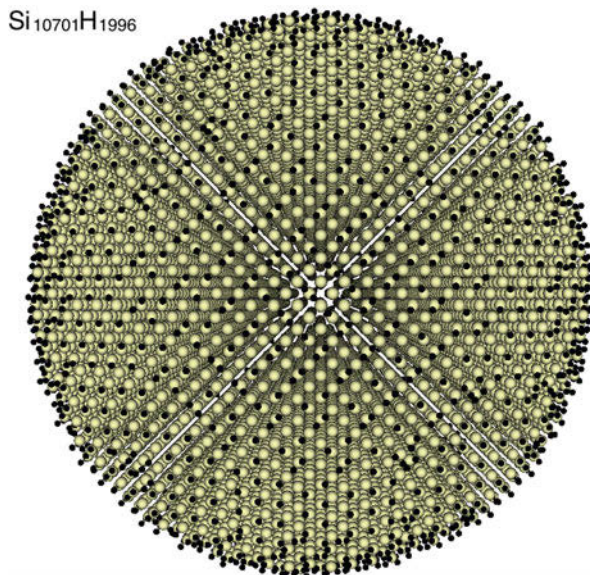
**Table 1** $Si_{525}H_{276}$, using 16 processors. The Hamiltonian dimension is 292584, where 1194 states need to be computed at each SCF step. The first step diagonalization by Chebyshev-Davidson costs 79755 #MVp and 221.05 CPU seconds; so the total #MVp spent on CheFS in CheFSI is 110,000. The polynomial degree used is $m = 17$ for Chebyshev-Davidson and $m = 8$ for CheFS. The first step diagonalization by TRLan requires 14909 #MVp and 265.75 CPU seconds

| Method | #MVp | #SCF steps | total_eV/atom | CPU(s) |
|--------|------|------------|---------------|--------|
| CheFSI | 189755 | 11 | −77.316873 | 542.43 |
| TRLan | 149418 | 10 | −77.316873 | 2755.49 |
| Diagla | 493612 | 10 | −77.316873 | 8751.24 |

symmetric operations of real-space pseudopotential methods, can routinely solve silicon clusters with several thousands of atoms.

We used the total energy per atom to assess accuracy of the final results. The number of iteration steps needed to reach self-consistency is given by `#SCF` . The number of matrix-vector products is given by `#MVp`.

In Table 1, we present a relatively small silicon cluster $Si_{525}H_{276}$, which is used to compare the performance of CheFSI with two eigenvector-based methods. The hardware for the computations illustrated here is modest, if not dated: an SGI Altix cluster consisting of 256 Intel Itanium processors at CPU rates of 1.6 GHz, sharing 512 GB of memory (but a single job is allowed to request at most 250 GB memory). Diagla and TRLan are standard eigensolver methods, which are not based on subspace filtering (Stathopoulos et al. 2000; Wu et al. 1999). Computational details such as grid spacings, pseudopotentials, and density functionals can be found in the literature (Zhou et al. 2006a, b).

We considered $Si_{2713}H_{828}$, $Si_{4001}H_{1012}$, and $Si_{9041}H_{1860}$. Results for these nanocrystals are show in Table 2. For these systems, the eigensolver Diagla became too slow to be practical. In the case of $Si_{2713}H_{828}$, we could still apply TRLan for the first step diagonalization for comparison, but we did not iterate until self-consistency was reached without excessive computational resources. Note that with the problem size increasing, Chebyshev-Davidson compares more favorably over TRLan. This is because we employed an additional trick in Chebyshev-Davidson, which corresponds to allowing the last few eigenvectors not to converge to the required accuracy. The number of the nonfully converged eigenvectors is bounded above by $act_{max}$, which is the maximum dimension of the active subspace. Typically $30 \leq act_{max} \leq 300$ for Hamiltonian size over a million where several thousand eigenvectors are to be computed. The implementation of this trick is rather straightforward since it corresponds to applying the CheFS method to the subspace spanned by the last few vectors in the basis that have not converged to required accuracy.

In Fig. 8, we illustrate the density of states for a large nanocrystal: $Si_{20389}H_{3076}$. This system presents a number of challenges. While the nanocrystal possesses some symmetry, the system is not periodic. There are 42,316 occupied states in this system. This requires the diagonalization in the Rayleigh-Ritz step of a large dense matrix.

We can compare the density of states to bulk silicon with the resulting energy distributions of energy levels for the nanocrystal, also shown in Fig. 8. The spectral features of this large nanocrystal have evolved to match the bulk. For this system, we can compute the complete evolution of the electronic structure from small molecular fragments of hydrogenated silicon nanocrystals to the bulk.

Within density functional theory, we can also examine the evolution of adding or removing electrons from a nanocrystal. First, we compute the energy to remove an electron and create a hole. For a semiconductor, this means we remove an electron from the highest occupied state and move it to the vacuum level. This energy is the ionization potential, $I$. The energy to create an electron in the lowest unoccupied state is called the electron affinity, $A$. These energies are defined as follows:

$$I = E(N - 1) - E(N)$$
$$A = E(N) - E(N + 1)$$

(16)

**Table 2** Performance of the CheFSI methods for large hydrogenated silicon nanocrystals. All calculations were done on 16 processors

| System | Dim. of $H$ | $n_{state}$ | #MVp | #SCF | total_eV/atom | 1st CPU | total CPU |
|--------|-------------|-------------|------|------|---------------|---------|-----------|
| $Si_{4001}H_{1012}$ | 1472440 | 8511 | 1652243 | 12 | −89.12338 | 18.63 h | 38.17 h |
| $Si_{6047}H_{1308}$ | 2144432 | 12751 | 2682749 | 14 | −91.34809 | 45.11 h | 101.02 h |
| $Si_{9041}H_{1012}$ | 2992832 | 19015 | 4804488 | 18 | −92.00412 | 102.12 h | 294.36 h |

where $N$ is the total number of electrons in the system and $E$ is the total electronic energy. In principle, the affinity and ionization energies are ground-state properties, and if the correct functional were known, these quantities would be accurately predicted by density functional theory.

We examine the scaling of the ionization potential and the electron affinity by assuming a simple power law behavior and fitting such a form to the calculated values (shown in Fig. 9):

$$I(D) = I_\infty + A/D^\alpha$$
$$A(D) = A_\infty + B/D^\beta$$

(17)

where $D$ is the diameter of the nanocrystal. A fit of these quantities results in $I_\infty = 4.5\,eV$, $A_\infty = 3.9\,eV$, $\alpha = 1.1$, and $\beta = 1.08$. The fit gives a quasiparticle gap of $E_{qp}(D \rightarrow \infty) = I_\infty - A_\infty = 0.6\,eV$ in the limit of an infinitely large dot. This value is in good agreement with the gap found for crystalline silicon using the local density approximation (Sham and Schlüter 1983).

The gap is not in good agreement with experiment owing to the failure of the local density approximation to describe band gaps of bulk semiconductors in general. We learned something interesting here. Even though $I$ and $A$ were computed from



**Fig. 8** Density of states for silicon. Top figure is for bulk silicon. The bottom figure is for a nanocrystal, $Si_{20389}H_{3076}$
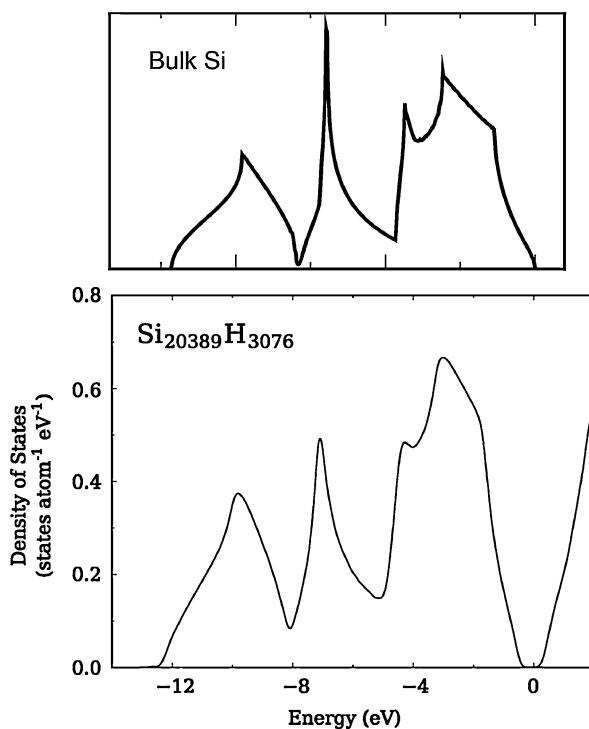
**Fig. 9** Evolution of the
ionization potential (IP) and
electron affinity (EA) with
quantum dot size. Also shown
are the eigenvalue levels for
the highest occupied
molecular orbital (HOMO)
and the lowest unoccupied
molecular orbital (LUMO)
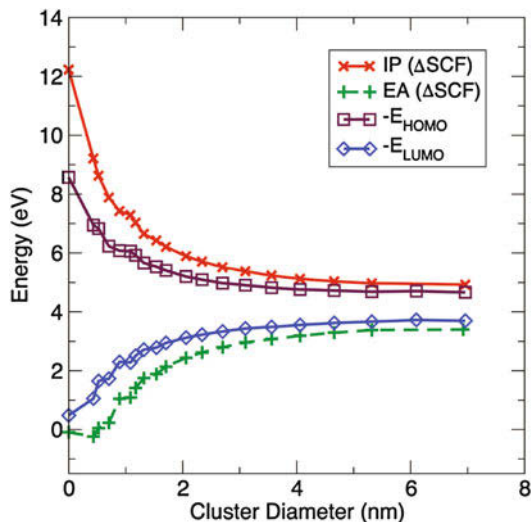


ground-state properties for large system, they do not yield correct properties for
the ground state. This situation reflects some fundamental flaw in our choice of
functionals and is not surprising as no known functional is "perfect."

A key aspect of this example is to show the scaling of the ionization potential and
electron affinity for nanocrystals ranging from silane ($SiH_4$) to hydrogenated silicon
systems containing thousands of atoms. We not only verify the limiting value of the
quasiparticle gap, but also we can ascertain how this limit is reached, i.e., how the
ionization potential and electron affinity scale with the size of the dot, and what the
relationship is between these quantities and the highest occupied and lowest empty
energy levels. In our example, we effectively have gone from the molecular limit
to the crystalline limit. We have spanned the entire nanocrystal regime, at least for
silicon.

Since values of $I$ and $A$ as calculated from density functional theory are reason-
ably accurate for atoms and molecules, one can ask how the size of the nanocrystal
affects the accuracy of such calculations. Unfortunately, experimental values for the
ionization potentials and electron affinities are not known for hydrogenated silicon
clusters and nanocrystals, a notable exception being silane, where the electron
affinity is negative, which means an electron added to the nanocrystal is unbound.

## 8    Conclusions

A pathway to improving algorithms is to focus on existing "bottlenecks" and remove
them. The resolution of these bottlenecks improves the performance of the code in
a systematic fashion.

A current bottleneck is the Rayleigh-Ritz step in Algorithm 4. This step involves
solving an eigenvalue problem for a dense matrix whose size is equal to the

number of eigenvalues. Suppose we wish to solve for a Si nanocrystal with 100,000 atoms or so. In such a situation, the Rayleigh-Ritz step would involve roughly a $200,000 \times 200,000$ dense matrix. With the algorithms outlined in this chapter coupled to current hardware, this is an insurmountable bottleneck.

A possible solution to this problem is to consider a combination of filters such that the eigenvalue problem is decomposed into a number of such problems each corresponding to a solution within an "energy window." Each window can then be handled separately, introducing an extra level of parallelism.

Implementing such an algorithm is difficult. The windows must be chosen so that each interval has approximately the same number of eigenvalues. This insures some reasonable load balancing. Often the eigenvalue spectrum is known for a similar system of interest, e.g., amorphous silicon or liquid silicon possesses occupied states that have a similar distribution as crystalline silicon. In this case, an approximate windowing is straightforward to implement without knowing the detailed eigenvalue spectrum.

A more challenging problem centers on not missing or duplicating states near the edge of an energy window. To avoid missing states, the energy windows should overlap. However, this overlap can result in double-counting states. Duplicates can be removed by single-value decomposition methods. Ongoing work is focused on this approach (Schofield et al. 2012a). The initial results are promising and should lead to our ability to do systems with thousands or tens of thousands of atoms.

Our focus in this chapter is on scalability and the extension of current methods to very large systems. We have not focused on an important related issue. Namely, can we also improve the accuracy of the computations? Local density functionals as those implemented in the present work are limited to weakly correlated systems, e.g., electronic materials like silicon, carbon, or gallium arsenide. Strongly correlated systems such as rare earths or rare earth oxides require more accurate functionals. These functionals often include nonlocal terms, e.g., the inclusion of Fock exchange. Such functionals can be implemented in real-space methods using the general framework outlined here (Boffi et al. 2016). The implementation of nonlocal functional results in a more complicated Hamiltonian and memory management issues. Typically, the scaling of such methods is notably limited.

Our quest to solve for large-scale systems should be apparent. Many problems require us to consider systems with very large numbers of atoms. For example, suppose we wish to examine a molecular system in an aqueous solution, a nucleation seed in the melt, or a solid-liquid interface. In these case, we might want to include thousands of atoms in a dynamical simulation to address important questions: How do materials dissolve? What is the structural and dynamical nature at a liquid-solid interface? Why do some nucleation sites promote crystal growth in only certain crystallographic directions? These systems are frequently beyond our reach. Numerical methods for these systems offer us the ability to test new models, gain new insights, and picture the microscope world in new ways.

# References

Alemany MMG, Jain M, Chelikowsky JR, Kronik L (2004) Real-space pseudopotential method for computing the electronic properties of periodic systems. Phys Rev B 69:075101

Alemany MG, Huang X, Tiago ML, Chelikowsky JR (2007) The role of quantum confinement in $p$-type doped indium phosphide nanowires. Nano Lett 7:1878

Andrade X, Aspuru-Guzik A (2013) Real-space density functional theory on graphical processing units: computational approach and comparison to gaussian basis set methods. J Chem Theor Comput 9:4360

Beck TL (2000) Real-space mesh techniques in density-functional theory. Rev Mod Phys 72:1041

Boffi NM, Jain M, Natan A (2016) Efficient computation of the Hartree-Fock exchange in real space with projection operators. J Chem Theory Comput 12:3614

Chan T-L, Chelikowsky JR (2010) Controlling lithium diffusion in semiconductor nanostructures by size and dimensionality. Nano Lett 10:821

Chan T-L, Tiago ML, Kaxiras E, Chelikowsky JR (2008) Size limits on doping phosphorus into silicon nanocrystals. Nano Lett 8:596

Chan T-L, Zayak AT, Dalpian GM, Chelikowsky JR (2009) Role of confinement on diffusion barriers in semiconductor nanocrystals. Phys Rev Lett 102:025901

Chan T-L, Lee AJ, Mok WK, Chelikowsky JR (2014) The interaction range of p-dopants in si [110] nanowires: determining the non-degenerate limit. Nano Lett 14:6306

Chelikowsky JR, Cohen ML (1992) Ab initio pseudopotentials for semiconductors. In: Landsberg PT (ed) Handbook on semiconductors, vol 1. Elsevier, Amsterdam, p 59

Chelikowsky JR, Troullier N, Saad Y (1994a) Finite-difference-pseudopotential method: electronic structure calculations without a basis. Phys Rev Lett 72:1240

Chelikowsky JR, Troullier N, Wu K, Saad Y (1994b) Higher-order finite-difference pseudopotential method: an application to diatomic molecules. Phys Rev B 50:11355

Chelikowsky JR, Chan T-L, Alemany MMG, Dalpian G (2011) Computational studies of doped nanostructures. Rep Prog Phys 74:046501

Dalpian G, Chelikowsky JR (2006) Self-purification in semiconductor nanocrystals. Phys Rev Lett 96:226802

Daniel J, Gragg WB, Kaufman L, Stewart GW (1976) Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization. Math Comput 30:772

Dirac PAM (1929) Quantum mechanics of many-electron systems. Proc R Soc A 123:714

Enkovaara J, Rostgaard C, Mortensen JJ, Chen J, Duak M, Ferrighi L, Gavnholt J, Glinsvad C, Haikola V, Hansen HA, Kristoffersen HH, Kuisma M, Larsen AH, Lehtovaara L, Ljungberg M, Lopez-Acevedo O, Moses PG, Ojanen J, Olsen T, Petzold V, Romero NA, Stausholm-Mller J, Strange M, Tritsaris GA, Vanin M, Walter M, Hammer B, Häkkinen H, Madsen11 GKH, Nieminen RM, Nørskov JK, Puska M, Rantala TT, Schiøtz J, Thygesen KS, Jacobsen KW (2010) Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. J Phys Condens Matter 22:253202

Fattebert J.-L., Bernholc J (2000) Towards grid-based O(N) density-functional theory methods: optimized nonorthogonal orbitals and multigrid acceleration. Phys Rev B 62:1713

Fornberg B, Sloan DM (1994) A review of pseudospectral methods for solving partial differential equations. Acta Numer 3:203

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP,

Smogunov A, Umari P, Wentzcovitch RM (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Goedecker S (1999) Linear scaling electronic structure methods. Rev Mod Phys 71:1085

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864

Kleinmann L, Bylander DM (1982) Efficacious form for model pseudopotentials. Phys Rev Lett 48:1425

Koch W, Holthausen MC (2000) A chemist's guide to density functional theory. Wiley-VCH, Weinheim

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133

Kresse G, Furthmülle J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169

Kronik L, Makmal A, Tiago ML, Alemany MMG, Jain M, Huang X, Saad Y, Chelikowsky JR (2006) PARSEC – the pseudopotential algorithm for real space electronic structure calculations: recent advances and novel applications to nano-structures. Phys Status Solidi B 243:1063

Lehoucq RB, Sorensen DC, Yang C (1998) ARPACK users guide: solution of large scale eigenvalue problems by implicitly restarted Arnoldi methods. SIAM, Philadelphia

Martin RM (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, Cambridge/New York

Ono T, Hirose K (2005) Real-space electronic-structure calculations with a time-saving double-grid technique. Phys Rev B 72:085115

Otsuka T, Miyazaki T, Ohno T, Bowler DR, Gillan MJ (2008) Accuracy of order-n density-functional theory calculations on dna systems using conquest. J Phys Condens Matter 20:29401

Parlett BN (1998) The symmetric eigenvalue problem. SIAM, Philadelphia

Phillips JC (1958) Energy-band interpolation scheme based on a pseudopotential. Phys Rev 112:685

Phillips JC, Kleinman L (1959) New method for calculating wave functions in crystals and molecules. Phys Rev 116:287

Saad Y (1992) Numerical methods for large eigenvalue problems. Wiley, New York

Saad Y, Chelikowsky JR, Shontz S (2010) Numerical methods for electronic structure calculations of materials. SIAM Rev 52:3

Sakai Y, Lee AJ, Chelikowsky JR (2016) First-principles non-contact atomic force microscopy image simulations with frozen density embedding theory. Nano Lett 16:3242

Schofield G, Chelikowsky JR, Saad Y (2012a) A spectrum slicing method for the Kohn-Sham problem. Comput Phys Commun 183:497

Schofield G, Chelikowsky JR, Saad Y (2012b) Using Chebyshev-filtered subspace iteration and windowing methods to solve the Kohn-Sham problem, chap. 6. In: Leszczynski J, Shulka MK (eds) Practical aspects of computational chemistry I: an overview of the last two decades and current trends, vol 167. Springer, Berlin

Seitsonen P, Puska MJ, Nieminen RM (1995) Real-space electronic-structure calculations: combination of the finite-difference and conjugate-gradient methods. Phys Rev B 51:14057

Sham LJ, Schlüter M (1983) Density functional theory of the energy gap. Phys Rev B 51:1888

Stathopoulos A, Öğüt S, Saad Y, Chelikowsky JR, Kim H (2000) Parallel methods and tools for predicting material properties. IEEE Comput Sci Eng 2:19

Teter MP, Allan DC, Arias TA, Joannopoulos JD (1992) Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. Rev Mod Phys 64:1045

Tiago ML, Zhou Y, Alemany MMG, Saad Y, Chelikowsky JR (2006) The evolution of magnetism in iron from the atom to the bulk. Phys Rev Lett 97:147201

Troullier N, Martins JL (1991) Efficient pseudopotentials for plane-wave calculations. Phys Rev B 43:1993

Wu K, Simon H (2000) Thick-restart Lanczos method for large symmetric eigenvalue problems. SIAM J Matrix Anal 22:602

Wu K, Canning A, Simon HD, Wang L-W (1999) Thick-restart Lanczos method for electronic structure calculations. J Comput Phys 154:156

Zhao Y, Du M-H, Kim Y-H, Zhang SB (2004) First-principles prediction of icosahedral quantum dots for tetravalent semiconductors. Phys Rev Lett 93:015502

Zhou Y, Saad Y (2007) A Chebyshev–Davidson algorithm for large symmetric eigenproblems. SIAM J Matrix Anal Appl 29:954

Zhou Y, Saad Y, Tiago M, Chelikowsky JR (2006a) Parallel self-consistent-field calculations via Chebyshev-filtered subspace acceleration. Phys Rev E 74:066704

Zhou Y, Tiago ML, Saad Y, Chelikowsky JR (2006b) Chebyshev-filtered subspace iteration method free of sparse diagonalization for solving the Kohn–Sham equation. J Comput Phys 219:172

# MP2- and RPA-Based Ab Initio Molecular Dynamics and Monte Carlo Sampling

# 25

Jürg Hutter, Jan Wilhelm, Vladimir V. Rybkin, Mauro Del Ben, and Joost VandeVondele

## Contents

### Abstract

Nonlocal correlation methods based on wave function theory are developed for application to condensed matter systems. These methods include MP2 and direct-RPA theory as well as double-hybrid functionals. Analytic gradients and stress tensors for MP2 theory in the gamma point approximation have been developed.

J. Hutter (✉) · J. Wilhelm · V. V. Rybkin
Institut für Chemie, Universität Zurich, Zurich, Switzerland
e-mail: hutter@chem.uzh.ch; jan.wilhelm@chem.uzh.ch; vladimir.rybkin@chem.uzh.ch

M. Del Ben
Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
e-mail: mdelben@lbl.gov

J. VandeVondele
Scientific Software and Libraries Unit, CSCS, ETH Zürich, Zürich, Switzerland
e-mail: joost.vandevondele@cscs.ch

523

Sampling complex systems at ambient temperature, for example, liquid water, becomes possible with efficient algorithms for massively parallel computers. Results show a qualitative improvement over standard local density functionals as well as hybrid functionals.

# 1    Introduction

Combining ensemble sampling techniques, like molecular dynamics or Monte Carlo methods, with first-principles electronic structure theory is commonly referred to as ab initio molecular dynamics (AIMD) (Marx and Hutter 2009). AIMD as a modern simulation technique was started by the seminal work of Roberto Car and Michele Parrinello (Car and Parrinello 1985). The Car-Parrinello (CP) method made AIMD simulations possible for a wide range of applications. In the early years, the CP method dominated the field and became a synonym for AIMD. The original CP method was proposed using Kohn-Sham density functional theory within the pseudopotential plane-wave framework. Standard simulation protocols for CP simulations were established and used by many groups. In recent years, new developments were able to overcome limitations of the CP method while keeping all of the benefits of the original method (Hutter 2012).

With the availability of more and more computer power, it became possible to also use Monte Carlo (MC) sampling techniques (McGrath et al. 2005a). In MC sampling the reuse of previous wave function information is much more difficult than in MD sampling. However, when atom-centered basis functions are used, this problem is considerably less severe than, for example, in plane-wave calculations. MC sampling has advantages over MD sampling in specific situations, like the simulation of the vapor-liquid coexistence curve (McGrath et al. 2005a, b), in situations where high barriers make sampling difficult (Schönherr et al. 2014) or if forces are not available (Del Ben et al. 2013b).

In the past, using other electronic structure methods than Kohn-Sham density functional theory was explored but never gained wide popularity. The lack of efficient implementations of correlated wave function methods for periodic systems and, especially, insufficient computational power to perform extensive sampling made such applications very demanding. A first step in the direction of the application of wave function correlation methods in AIMD was the successful incorporation of efficient exact exchange functionals in condensed matter simulations (Todorova et al. 2006; Guidon et al. 2008). Only in recent years, it has been possible to combine extensive sampling approaches with wave function correlation methods (Del Ben et al. 2013b, 2015b). In this article we will summarize developments of electronic structure methods that made the application of wave function correlation methods in AIMD possible. We will also investigate the results from pioneering applications.

## 2 MP2 and RPA Correlation Energy Methods

The generalized gradient model (GGA) within Kohn-Sham DFT had tremendous success in numerous applications. Yet, the model has significant shortcomings that also influence, e.g., the quality of simulations for aqueous chemistry. Most notable are the absence of van der Waals interactions and a significant self-interaction error. The former leads, e.g., to an underestimation of the water density (Schmidt et al. 2009; Wang et al. 2011), while the latter leads to an underestimated bandgap, with implications for redox chemistry (Adriaanse et al. 2012) and the static dielectric constant (Schönherr et al. 2014). It is possible to go beyond GGA in various ways, and improvements have been made by including various descriptors of the electronic system to yield models with improved accuracy. In an attempt to classify this progress, Perdew et al. (2005) employed the metaphor of a "Jacob's ladder" for which each rung of the ladder introduces more descriptors of the electronic system and yields models with improved accuracy. This ladder has currently five rungs which include as descriptors (1) the electronic density; (2) its gradient; (3) the kinetic energy density; (4) the occupied molecular orbitals (MO), usually in the form of Hartree-Fock exchange; and (5) the unoccupied or virtual MOs. GGA belongs to the second rung, while hybrid functionals belong to the fourth rung. At the fifth rung, the inclusion of the virtual orbitals allows for taking into account nonlocal, dynamical, electron correlation effects that contribute to the long-range van der Waals (vdW) dispersion interactions. Many of the various functionals on the fifth rung are based on either the random phase approximation (RPA) (Eshuis et al. 2012; Paier et al. 2012; Grimme and Steinmetz 2016) or second-order Møller-Plesset perturbation theory (MP2) (Grimme 2006; Goerigk and Grimme 2011), in the form of double hybrids (DH). Direct use of RPA and MP2 energies and variations thereof is also of interest for many applications. As a basic building block, the calculation of the RPA and MP2 energies has to be provided. We will in the subsequent sections concentrate on the theory and implementation of these methods.

### 2.1 Theory

We present briefly the theoretical framework of the methods and refer to the original works for more details. First, the resolution of the identity approximation for two-electron repulsion integrals (ERIs) is introduced, and then its application to the different correlation methods is formulated. The following index notation has been adopted: $i, j, k, \ldots$ refer to canonical occupied molecular orbitals (MOs); $a, b, c, \ldots$ to canonical virtual MOs; $\mu, v, \lambda, \ldots$ to atomic orbital basis set functions (AO); and $P, Q, R, \ldots$ to auxiliary basis set functions (AUX). The one-electron MO and AO functions are symbolized with $\psi$ and $\phi$, respectively. The number of occupied and virtual orbitals is denoted by $o$ and $v$ while the total number of primary and auxiliary basis functions as $n$ and $N_a$. In order to express, in general, the system size, the symbol $N$ is used.

### 2.1.1 Periodic Boundary Conditions

In order to minimize boundary effects, simulations of condensed matter systems are performed using periodic boundary conditions. In periodic systems the infinite number of particles to be treated is reduced to the unique particles in the simulation cell, and the calculation of the electronic energy involves an integration over the first Brillouin zone. For disordered systems, as, for example, liquids, this integration can be replaced by a single point in the center of the integration volume, the $\Gamma$-point. In the following chapters, we will always assume that this $\Gamma$-point approximation is invoked, and we can further assume that the single-particle wave functions at this point are real.

### 2.1.2 The Resolution of the Identity (RI) Approximation

The two-electron repulsion integrals (ERI), in Mulliken notation, of the type $(ia|jb)$ are of central importance for all the methods presented. Within the RI approximation (Whitten 1973; Dunlap et al. 1979), based on the Coulomb metric (Vahtras et al. 1993), these integrals are factorized according to:

$$(ia|jb)_{RI} = \sum_{PQ} (ia|P)(P|Q)^{-1}(Q|jb) \tag{1}$$

where $(P|Q)^{-1}$ is the inverse matrix of the Coulomb metric,

$$(P|Q) = \iint \phi_P(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_Q(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \tag{2}$$

The auxiliary basis set size $N_a$ grows only linearly with the system size (Feyereisen et al. 1993). The main advantage of the RI approximation is that four-center integrals of the type $(ia|jb)$ are computed from three- and two-center ERIs. This allows to strongly reduce the effort for the integral computation without significant loss of accuracy (Weigend and Häser 1997). Since the $(P|Q)$ matrix is positive definite, its inverse can be efficiently obtained from a Cholesky decomposition

$$(P|Q) = \sum_R L_{PR} L_{RQ}^T \tag{3}$$

followed by inversion of the triangular matrix $\mathbf{L}$. In this way the factorization of the $(ia|jb)$ integrals can be expressed in a compact form as:

$$(ia|jb)_{RI} = \sum_P B_P^{ia} B_P^{jb}. \tag{4}$$

$\mathbf{B}$ is a matrix with $ov$ rows and $N_a$ columns, given by:

$$B_P^{ia} = \sum_R (ia|R) L_{PR}^{-1}. \tag{5}$$

Since the three-center integrals $(ia|R)$ are computed starting from integrals over AOs $(\mu\nu|R)$, the final expression for the $B_P^{ia}$ elements reads:

$$(ia|P) = \sum_v C_{va} \sum_\mu C_{\mu i} \sum_R (\mu\nu|R) L_{PR}^{-1} \tag{6}$$

where $\mathbf{C}$ is the MO coefficient matrix.

Computation of the $\mathbf{B}$ matrix thus includes, first, calculation of $(P|Q)$ and via Cholesky decomposition and triangular inversion, $\mathbf{L}^{-1}$. Second, three-center integrals $(\mu\nu|R)$ are computed and transformed using the $\mathbf{C}$ and $\mathbf{L}^{-1}$ matrices (Eq. 6). The first two steps formally scale $O(N^2)$ and $O(N^3)$, respectively, while three-center integral computation requires formally $O(N^3)$ operations and integral transformations scale $O(N^4)$. Within the RI approximation, the asymptotically dominating step in computing $\mathbf{B}$ is thus the index transformation.

### 2.1.3 RI-MP2 Method

In second-order Møller-Plesset perturbation theory (Møller and Plesset 1934), the correlation energy $E^{(2)}$ for a closed-shell system is given by:

$$E^{(2)} = -\sum_{i \leq j}^o (2 - \delta_{ij}) \sum_{ab}^v \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}. \tag{7}$$

where $\varepsilon_a$ and $\varepsilon_i$ are orbital energies. In a canonical MP2 energy algorithm, the time-limiting step is the computation of the $(ia|jb)$ integrals obtained from the ERIs over AO $(\mu\nu|\lambda\sigma)$ via four consecutive integral transformations. The computational effort for the first quarter transformation (no sparsity considered) is $O(on^4)$, making the MP2 energy calculation a method scaling as $O(N^5)$. The application of the RI approximation to MP2 is straightforward (Feyereisen et al. 1993) and consists in replacing $(ia|jb)$ integrals with the approximate $(ia|jb)_{RI}$ given in Eq. 4. The computation of $(ia|jb)_{RI}$ requires $O(o^2v^2N_a)$ operations implying that the RI-MP2 method is also scaling $O(N^5)$. The main reason for the speedup observed in RI-MP2 lies in the strongly reduced integral computation part.

### 2.1.4 RI Direct Random Phase Approximation Correlation Energy Method

The RPA correlation energy is given as the difference between the zero-point energy of two harmonic oscillator excitation problems for which the first includes a correlated ground state (RPA) and the second does not (configuration interaction singles) (Furche 2008; Scuseria et al. 2008). Within the direct-RPA (dRPA) approach, which is RPA without exchange contributions (Eshuis et al. 2010), $E_c^{\text{dRPA}}$ can be expressed in terms of a frequency integral

$$E_c^{\text{dRPA}} = \frac{1}{2} \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \text{Tr}(\ln(\mathbf{1} + \mathbf{Q}(\omega)) - \mathbf{Q}(\omega)), \tag{8}$$

with the frequency-dependent matrix $\mathbf{Q}(\omega)$ in the RI basis which is determined by

$$Q_{PQ}(\omega) = 2 \sum_i^o \sum_a^v B_P^{ia} \frac{\varepsilon_a - \varepsilon_i}{\omega^2 + (\varepsilon_a - \varepsilon_i)^2} B_Q^{ia}. \tag{9}$$

For a given $\omega$, the computation of the integrand function in Eq. 8 using Eq. 9 requires $O(N^4)$ operations. The integral of Eq. 8 can be efficiently calculated by Clenshaw-Curtis numerical quadrature, and usually $30-40$ quadrature points are enough to achieve micro-Hartree accuracy. Thus, the introduction of the resolution of the identity approximation and the frequency integration techniques for computing $E_c^{\mathrm{dRPA}}$ lead to a reduction of the computational cost to $O(N^4 N_q)$ and $O(N^3)$ storage, where $N_q$ is the number of quadrature points.

## 2.2    Nuclear Gradients and Stress Tensor

Calculation of analytic derivatives is of central importance for applications in electronic structure theory. The specific derivatives needed for nuclear gradients and stress tensor have received most attention as they are connected to basic structural properties of molecules and materials. In almost all applications, the availability of analytic derivatives is of advantage for reasons of accuracy and efficiency. The quantum chemistry community has a long tradition in developing algorithms for analytic derivatives, especially also for non-variational methods as the ones considered here (Handy and Schaefer 1984). These techniques have been used to derive derivatives for double-hybrid functionals (Neese et al. 2007) and RPA methods (Burow et al. 2014). For periodic systems, derivatives for MP2 in the restricted (Del Ben et al. 2015a) and unrestricted (Rybkin and VandeVondele 2016), as well as RPA (Ramberger et al. 2017) methods, have been reported.

### 2.2.1    Analytic Derivatives for RI-MP2

The analytic derivative of the RI-MP2 energy $E_{\mathrm{RI}}^{(2)}$ with respect to a perturbation parameter $x$, for a closed-shell restricted Hartree-Fock wave function, is given by Weigend and Häser (1997):

$$E_{\mathrm{RI}}^{(2)x} = \frac{dE_{\mathrm{RI}}^{(2)}}{dx} = 4 \sum_Q^{\mathrm{AUX}} \sum_{\mu\nu}^{\mathrm{AO}} \Gamma_{\mu\nu}^Q (\mu\nu|Q)^x - 2 \sum_{PQ}^{\mathrm{AUX}} \Gamma^{PQ} (P|Q)^x$$

$$+ 2 \sum_{pq}^{\mathrm{MO}} \left[ P_{pq}^{(2)} F_{pq}^{(x)} - W_{pq}^{(2)} S_{pq}^{(x)} \right]. \tag{10}$$

In the above expression, for each summation, a common structure can be recognized, that is, the contraction of terms involving AO derivatives $(\mu\nu|Q)^x$, $(P|Q)^x$, $F_{pq}^{(x)}$, $S_{pq}^{(x)}$, with elements of the intermediates $\Gamma_{\mu\nu}^Q$, $\Gamma^{PQ}$, $P_{pq}^{(2)}$, $W_{pq}^{(2)}$. The

first two summations involve the contraction of three- and two-center RI integral derivatives $(\mu\nu|Q)^x$, $(P|Q)^x$ with corrections to the two-particle density matrix (2-PDM), $\Gamma^Q_{\mu\nu}$ and $\Gamma^{PQ}$. These quantities are available from the energy calculation (Del Ben et al. 2015a). The last summation in Eq. 10 consists in the contraction of $P^{(2)}_{pq}$, the MP2 correction to the 1-PDM, and $W^{(2)}_{pq}$, the MP2 correction to the energy-weighted density matrix, with the skeleton derivatives of the Fock and overlap matrix elements

$$F^{(x)}_{pq} = \sum_{\mu\nu}^{\text{AO}} C_{\mu p} \left[ h^x_{\mu\nu} + \sum_{\lambda\sigma} P^{\text{HF}}_{\lambda\sigma}(\mu\nu|\lambda\sigma)^x - \frac{1}{2}\sum_{\lambda\sigma} P^{\text{HF}}_{\lambda\sigma}(\mu\lambda|\nu\sigma)^x \right] C_{vq} \qquad (11)$$

$$S^{(x)}_{pq} = \sum_{\mu\nu}^{\text{AO}} C_{\mu p} S^x_{\mu\nu} C_{vq}. \qquad (12)$$

In Eq. 11, $h^x_{\mu\nu}$ and $(\mu\nu|\lambda\sigma)^x$ are the derivatives of the one-electron Hamiltonian integrals and the ERIs, while $P^{\text{HF}}_{\mu\nu} = 2\sum_i^o C_{\mu i} C_{vi}$ is the Hartree-Fock density matrix. In order to take advantage from sparsity, the update of the $E^{(2)}_{\text{RI}}$ derivative is performed in the AO basis, after back transformation of $P^{(2)}_{pq}$ and $W^{(2)}_{pq}$ from the MO basis.

The diagonal blocks of the 1-PDM $P^{(2)}_{pq}$ can be calculated again using quantities from the energy calculation. The virtual occupied block of $P^{(2)}$ contains information related to the orbital relaxation caused by the perturbation $x$, i.e., first-order response of the MO coefficients. It is computed as the solution of the Z-vector equations (Handy and Schaefer 1984)

$$\sum_a^v \sum_i^o \left[ \delta_{ij}\delta_{ab}(\varepsilon_a - \varepsilon_i) + A_{aibj} \right] P^{(2)}_{ai} = -L_{bj} \qquad (13)$$

where $A_{aibj}$ is an element of the orbital Hessian matrix

$$A_{aibj} = 4(ai|bj) - (ab|ij) - (aj|bi), \qquad (14)$$

and $L$ is a specific RI-MP2 Lagrangian matrix given by:

$$L_{bj} = 2\sum_a^v \sum_Q^{\text{AUX}} (ba|Q)\Gamma^Q_{ja} - 2\sum_i^o \sum_Q^{\text{AUX}} (ij|Q)\Gamma^Q_{ib}$$

$$+ \sum_{ac}^v P^{(2)}_{ac} A_{acbj} + \sum_{ik}^o P^{(2)}_{ik} A_{ikbj}. \qquad (15)$$

Due to the large size of the orbital Hessian matrix $\mathbf{A}$ ($ov \times ov$), the linear system of Eq. 13 is commonly solved by iterative techniques. Rather than calculating and storing the full $\mathbf{A}$, at each iteration, the matrix-vector product $\sum_{ia} X_{ai} A_{aibj}$ is computed, with $\mathbf{X}$ being a trial solution. Finally, the MP2 correction to the energy-weighted density matrix $W_{pq}^{(2)}$ can be calculated from the relaxed 1-PDM $P_{pq}^{(2)}$. The methodology presented is of general validity for any perturbation parameter $x$. In particular, for the calculation of the forces acting on the ions, the gradients of $E_{\mathrm{RI}}^{(2)}$ with respect to the atomic positions have to be computed.

The RI-MP2 contribution to the total stress tensor is calculated according to Nielsen and Martin (1985) and Del Ben et al. (2015a):

$$\Pi_{\alpha\beta}^{(2)} = -\frac{1}{3V} \sum_{\gamma=1}^{3} \frac{\partial E_{\mathrm{RI}}^{(2)}}{\partial h_{\alpha\gamma}} h_{\gamma\beta}^{T} \tag{16}$$

where $h_{\alpha\gamma}$ are elements of the matrix of the cell vectors (Bravais lattice vectors) given by $a_1$, $a_2$, and $a_3$, that is, $\mathbf{h} = [a_1, a_2, a_3]$.

## 2.3 Implementation and Performance

### 2.3.1 RI Gaussian and Plane-Wave Method

The Gaussian and plane-wave (GPW) method has been shown to be an efficient approach for computing ERIs especially when periodic boundary conditions are considered (Del Ben et al. 2012). The basis of the GPW approach for the computation of the ERIs is the direct formulation of the half-transformed integrals of the type $(ia|\lambda\sigma)$ in terms of the electrostatic potential $v^{ia}$ of the pair density $\rho^{ia}$

$$
\begin{aligned}
(ia|\lambda\sigma) &= \iint \psi_i(\mathbf{r}_1)\psi_a(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \\
&= \int \left[ \int \frac{\rho^{ia}(\mathbf{r}_1)}{\mathbf{r}_{12}}d\mathbf{r}_1 \right] \phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)d\mathbf{r}_2 \\
&= \int v^{ia}(\mathbf{r}_2)\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)d\mathbf{r}_2.
\end{aligned}
\tag{17}
$$

The form of the last equation is essentially identical to the one used in the GPW method (Lippert et al. 1997) to compute matrix elements of the Hartree potential. Thus, the highly efficient implementation of that operation in the CP2K code can be directly used, and we refer to VandeVondele et al. (2005a) for a detailed discussion.

Within the RI approximation, two types of ERIs have to be computed, the two-center $(P|Q)$ and three-center $(ia|P)$ integrals. Three-center integrals are computed starting from the integrals over AOs that are subsequently transformed with $\mathbf{C}$ and $\mathbf{L}^{-1}$. The index transformation over the auxiliary basis can be avoided,

since it is possible to directly compute half-transformed integrals for an associated density $\rho$ as

$$
\begin{aligned}
(\mu v|P) &= \sum_R (\mu v|R)L_{PR}^{-1} = \iint \phi_\mu(\mathbf{r}_1)\phi_v(\mathbf{r}_1)\frac{1}{\mathbf{r}_{12}}\left[\sum_R \phi_R(\mathbf{r}_2)L_{PR}^{-1}\right]d\mathbf{r}_1 d\mathbf{r}_2 \\
&= \int \phi_\mu(\mathbf{r}_1)\phi_v(\mathbf{r}_1)\left[\int \frac{\rho^P(\mathbf{r}_2)}{\mathbf{r}_{12}}d\mathbf{r}_2\right]d\mathbf{r}_1 = \int \phi_\mu(\mathbf{r}_1)\phi_v(\mathbf{r}_1)v^P(\mathbf{r}_1)d\mathbf{r}_1.
\end{aligned}
\tag{18}
$$

The same approach holds for the $(P|Q)$ integrals. An alternative way to calculate these integrals over Gaussian basis functions using analytic Ewald techniques is also available (Wilhelm et al. 2016).

Of central importance in GPW is the representation of the density on a regular grid, which is equivalent to an expansion of the density in an auxiliary basis of plane waves (PW). The expansion is given by

$$
\rho^P(\mathbf{r}) \approx \frac{1}{\Omega}\sum_{|\mathbf{G}|\leq G_c}\rho^P(\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{r}}
\tag{19}
$$

where the sum over the reciprocal lattice vectors $\mathbf{G}$ is determined by the resolution of the grid. $\rho^P(\mathbf{G})$ are the Fourier coefficients of the density, and $\Omega$ is the volume of the simulation cell. Fast Fourier transforms (FFTs) efficiently change between real-space and reciprocal-space representations. In reciprocal space, it becomes straightforward to solve the Poisson equation for the potential $v^P$

$$
v^P(\mathbf{G}) = \frac{4\pi}{G^2}\rho^P(\mathbf{G})
\tag{20}
$$

and an additional back FFT will yield the potential in real space. For $\mathbf{G} = 0$ the value of the potential is set to zero, corresponding to a constant shift of the potential, enforcing zero average. Due to the orthogonality of the occupied-virtual orbitals, this shift has no influence on the final value of the $(ia|P)$ integrals. Within the GPW method, pseudopotentials have to be employed in order to have sufficiently smooth densities. Once the potential $v^P$ is available, the numerical integration over the basis functions is performed by summing over the grid points. For a given $|P)$, all matrix elements that are non-zero within a given threshold can be obtained in linear scaling time. A further gain in efficiency is obtained by employing a multigrid technique that represents the potential $v^P$ on grids with increasingly coarser grid spacing. Finally, $(\mu v|P)$ integrals are transformed into MO ERIs using (sparse) matrix multiplication as implemented in the DBCSR library (Borštnik et al. 2014).

The accuracy of the RI approximation is directly related to the completeness of the auxiliary basis set. It has been shown that it is possible to generate compact auxiliary basis sets with a size of approximately three times the size of the orbital

**Table 1** Counterpoise-corrected PWPB95-D3 interaction energies (kcal/mol) of the S22 set obtained employing the cc-TZ, cc-QZ, and aug-cc-TZ basis sets (Del Ben et al. 2013a). Mean absolute deviation (MAD), root mean square deviation (RMSD), maximum deviation (Max), and mean absolute percentage deviation (MA%D) of each basis set with respect to the reference values (Goerigk and Grimme 2011) are reported

|        | cc-TZ | cc-QZ | aug-cc-TZ |
|--------|-------|-------|-----------|
| MAD    | 0.43  | 0.27  | 0.15      |
| RMSD   | 0.52  | 0.31  | 0.21      |
| Max    | 1.24  | 0.71  | 0.59      |
| MA%D   | 7.45  | 5.42  | 3.48      |

basis set with a transferable overall accuracy of $\approx 1$mHartree/atom (Weigend et al. 1998). Specific RI basis sets for the application in condensed systems have been generated and tested (Del Ben et al. 2013a).

The setup used in the applications presented later has been tested with respect to the basis set error for correlated calculations. The slow convergence of correlated methods with basis set size is a well-studied problem in quantum chemistry. In Table 1 results for three basis sets are presented for calculations of the S22 reference sets using a double-hybrid functional. An improvement of results with the increase of the basis set can be observed. However, it should be noted that the reference used is itself not fully converged and that there are other parameters, e.g., the pseudopotential used, that result in differences. It can be seen that the results obtained with the cc-TZ basis, that is used in the liquid water simulations, is of good quality but cannot be considered fully converged.

### 2.3.2 Parallel Implementation of the RI-GPW Methods

In this section the parallelization strategies for the methods introduced are presented. The algorithms are split in two steps: the first deals with the computation of the ERIs $(ia|P)$ and is common for all methods, and the second is specific to the type of correlation energy calculated. The parallelization is achieved with a multi-level hybrid OpenMP/MPI scheme using a careful process layout. The first level of parallelization corresponds to distributing the work performed for a single given auxiliary basis function $\phi_P$ or vector $|P) = \sum_R \phi_R L_{PR}^{-1}$. The second level of parallelization corresponds to a distribution of these nearly independent calculations. The $N_p$ processes available in total are split in $N_G$ groups, each group working on a given $\phi_P$ or $|P)$ and each consisting of $N_w$ processes ($N_p = N_G N_w$). The first level of parallelization involves parallel FFTs, halo-exchanges, and sparse matrix multiplications over $N_w$ processes and corresponds to the standard parallelization scheme for DFT calculations in CP2K (VandeVondele et al. 2005a). The second level is straightforward and only requires intergroup redistribution of two-center ERIs $(Q|P)$ in order to calculate $\mathbf{L}^{-1}$.

The total workload for the integral computation is distributed by splitting the auxiliary basis functions into $N_G$ ranges. Additionally, each of the $N_w$ processes within a group is assigned a range of virtual orbital indices, while the occupied

orbital index is fully replicated. Once the integrals $(ia|P)$ are available, we can proceed to calculate the RI-MP2 energy. Since $N_w$ is usually small compared to the total number of processes, the virtual index $a$ is distributed over a small number of MPI tasks within the group $G$, while the auxiliary index $P$ is distributed over a large number of $N_G$ groups. First, the independent $ij$ pairs ($i \leq j$) are distributed over the $N_G$ groups. For each $ij$ pair, the full range of the auxiliary index $P$ is collected on a local buffer from all other groups while keeping the virtual index distribution within the group. The $(ia|jb)$ integrals are generated for the actual $ij$ pair in a matrix-multiplication fashion (Eq. 4) requiring only a small amount of communication within the group. Once the $(ia|jb)$ are available, they are accumulated into the MP2 energy according to Eq. 7, requiring an additional negligible amount of communication within the group. With this choice, the main source of intergroup communication in the parallel algorithm is related to the redistribution of the $B_{ia}^P$ integrals, required for each $ij$ pair. The time-determining step is the $(ia|jb)$ integral generation that is essentially a local matrix multiplication. This allows to fully exploit the performance of highly optimized routines and can be further accelerated by employing a hybrid implementation that utilizes graphics processing units (GPUs).

The dRPA correlation energy implementation is based on the method developed by Del Ben et al. (2013a). The parallel algorithm has a two-level workload distribution. The first level corresponds to the distribution of the work necessary for a given quadrature point of the integral in Eq. 8. The second level distributes the calculation of the independent quadrature points over subgroups of processes. The **B** matrix has to be replicated within the groups in order for the algorithm to proceed independently for each integration point. As a first task, the matrix **B**′ is calculated as $\mathbf{G}(\omega)\mathbf{B}$. Since $\mathbf{G}(\omega)$ is a $ov \times ov$ diagonal matrix with elements $G_{ia,ia}(\omega) = (\varepsilon_a - \varepsilon_i)((\varepsilon_a - \varepsilon_i)^2 + \omega^2)^{-1}$, the calculation of **B**′ proceeds without communication. The time-determining step of the algorithm is the calculation of the matrix $\mathbf{Q}(\omega)$ computed as $2\mathbf{B}^\mathbf{T}\mathbf{B}'(\omega)$. This task is performed as a standard parallel matrix multiplication. The calculation of $\mathrm{Tr}[\ln(\mathbf{Q}(\omega) + \mathbf{1})]$ can be efficiently carried out by considering the identity $\mathrm{Tr}[\ln \mathbf{A}] = \ln(\mathrm{Det}[\mathbf{A}])$, that is:

$$\mathrm{Tr}[\ln(\mathbf{Q}(\boldsymbol{\omega}) + \mathbf{1})] = 2\sum_{i=1}^{N_a} \ln(U_{ii}) \qquad (21)$$

where the **U** matrix is the Cholesky decomposition of $\mathbf{Q}(\omega) + \mathbf{1}$.

The excellent performance of the presented implementation can be seen from the results in Fig. 1. Panel (a) shows the achieved scaling for a 64-water-molecule system using a range of 512 to 32,768 processes. Both methods, MP2 and dRPA, show similar overall scalability with $\approx$80% efficiency over the full range tested. In panel (b) timings for the RI integral generation as well as the energy calculation for different methods are shown. In a doubly logarithmic plot, the system size (generated as increasing water systems) is given versus the total CPU time. All measured timings show a linear dependency, resulting in an estimate for the true
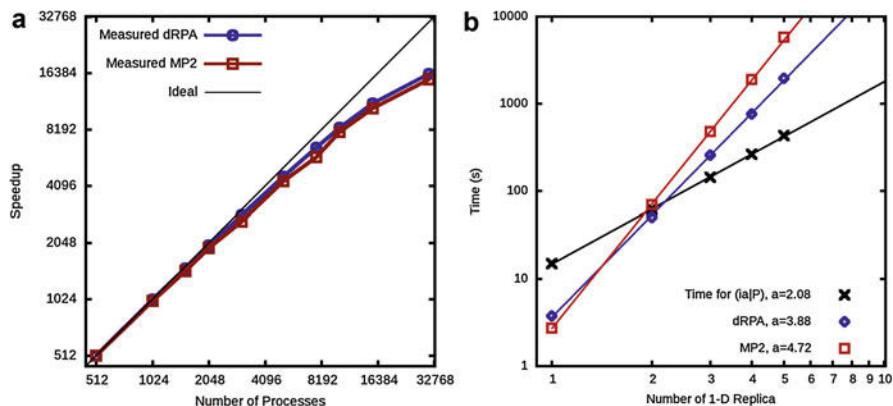
**Fig. 1** (**a**) Measured speedup with respect to 512 processes for the calculation of the MP2 and dRPA energy of 64 bulk water molecules. dRPA calculation performed employing 60 quadrature points for the numerical integration. (**b**) CPU time as a function of the number of replicas of the supercell containing 32 $H_2O$ molecules. Twenty quadrature points were used for dRPA. Lines represent a linear two-parameter fit of the form $y = bx^a$

scaling of the method from the observed slope of the linear fit. The observed scaling laws closely reproduce the theoretical scaling from the analysis of the implemented algorithms.

## 3　　Application: Properties of Water

Despite the apparent simplicity of the $H_2O$ molecule, bulk liquid water has a rich chemistry and physics. This is commonly illustrated referring to its many anomalous properties. These macroscopic properties are well characterized experimentally, but resolving the underlying atomistic picture is far more difficult (Errington and Debenedetti 2001), and even the structure of the neat liquid is still debated intensely (Nilsson and Pettersson 2011; Soper 2013). Ultimately, our understanding must go beyond the neat liquid. The properties of water, as a solvent or reactant, in the bulk, near interfaces or in confinement are of central importance in many fields, such as biology, electrochemistry, catalysis, earth, and climate science.

　　Theory and simulations can complement the experimental efforts and have a long tradition. For example, liquid water was among the first systems studied when molecular dynamics (MD) based on empirical potentials became available in the 1970s (Rahman and Stillinger 1971). Recent models are mostly derived based on high-level electronic structure calculations and show excellent agreement with the experiment for a variety of properties at low computational cost. However, empirical methods might fail when applied outside their fitting range, and the effort to parameterize the models and refine the employed functional forms cannot be underestimated. This challenge rises considerably as soon as solutes come into

play or if more complicated chemical phenomena, such as autodissociation and reactivity, must be taken into account. Computing the intermolecular interactions during the entire simulation using electronic structure theory is an alternative to empirical models as complexity increases, since no assumptions on the form of the interactions must be made. The pioneering application of electronic structure theory to bulk liquid water was based on Kohn-Sham density functional theory using a GGA density functional approximation (Laasonen et al. 1993), and the same class of functionals has been employed for most subsequent applications (see, e.g., VandeVondele et al. 2005b; McGrath et al. 2005a). Recently, also hybrid functionals have been used for the simulation of liquid water (Todorova et al. 2006; Guidon et al. 2008; DiStasio et al. 2014). The importance of vdW interactions for the description of liquid water has also become apparent, and current applications rely on different available approaches to augment standard local functionals. The two main approaches, empirical pair potentials (Grimme et al. 2010) and explicit nonlocal correlation functionals (Dion et al. 2004), are outside the Perdew classification. The performance of these approaches has been extensively tested, and results show a systematic improvement upon the uncorrected GGA or hybrid functionals (Schmidt et al. 2009; DiStasio et al. 2014).

We assess the performance of various methods, with a focus on MP2, RPA, and double-hybrid functionals. In particular, the structural and dynamical properties of bulk liquid water have been studied by means of Monte Carlo (MC) and molecular dynamics (MD) simulations (Del Ben et al. 2013b, 2015b). The MC simulations have been performed in the NpT ensemble under ambient pressure and temperature and are focused on the structural properties, while MD has been employed to obtain dynamical observables, namely, the infrared spectrum and the diffusion constant (Del Ben et al. 2015b). A significant advantage of the MC scheme is that the sampling of the phase space is solely determined by a total energy-based criterion and that forces and stresses are not explicitly required, simplifying the implementation. The downside is that dynamical properties are not available and that suitable moves are needed to efficiently explore large portions of the configuration space.

## 3.1   Computational Setup

All calculations presented have been performed with the CP2K program. The CP2K code makes use of a dual representation for the electronic density and MOs in terms of Gaussian and plane-wave (GPW) (Lippert et al. 1997; VandeVondele et al. 2005a). Unless specified otherwise, the Gaussian basis is of correlation-consistent triple-zeta quality (Dunning 1989; Del Ben et al. 2012), while the PW cutoff is set to 800 Ry. To efficiently expand the density and orbitals in plane waves within GPW, core electrons are replaced by pseudopotentials (Goedecker et al. 1996) that have been parametrized for the employed functionals. The exact exchange calculations are performed employing a $\Gamma$-point implementation making use of a truncated Coulomb operator to avoid divergence of the energy (Guidon

et al. 2008, 2009); the truncation radius is set to 6 Å. The RI auxiliary basis is specifically fitted for this purpose (Del Ben et al. 2015b). All dRPA calculations have been performed employing KS Perdew-Burke-Ernzerhof (PBE) (Perdew et al. 1996) orbitals as input. More details on the computational setups are reported in supporting information of Del Ben et al. (2015b). The model system is made of 64 $H_2O$ molecules in a cubic box under periodic boundary conditions (PBC). The Monte Carlo simulations have been performed with thermodynamic constraints set to ambient conditions, that is, $T = 295$ K and $p = 1$ bar. The MC efficiency is improved with the presampling of moves (Iftimie et al. 2000); in the actual case, the approximated potential is calculated at the DFT level employing the PBE1W (García-González et al. 2007) functional, for which the basis and D3 parameters have been specifically refitted in order to closely approximate the energy of the wave function methods and thus to increase acceptance of moves. Molecular dynamics (MD) simulations which started from equilibrated MC configurations and used a multiple time-step scheme are previously proposed in the context of hybrid density functionals (Guidon et al. 2008).

## 3.2    Structural Properties of Liquid Water

In Fig. 2 the density fluctuation over the MC trajectory obtained with the MP2 and RPA methods is reported. At the RPA level, the density quickly equilibrates, giving an average density of 0.994 g/mL in excellent agreement with the experiment. The root mean square deviation of the instantaneous density is 0.015 g/mL similar to that of the MP2 method which has an average density of 1.020 g/mL. As RPA is computationally less demanding than MP2, roughly twice the number of MC cycles has been performed relative to MP2 (26.8 kcycles vs. 14.6 kcycles). Nevertheless, finite sampling times cannot exclude sudden structural rearrangements on longer timescales. This was observed at the PWPB95-D3 level (Del Ben et al. 2015b), where, initially, the fluctuations of the instantaneous density seem to equilibrate to a stable average value of 1.002 g/mL. However, later a sharp transition occurred to a high-density phase, which was not fully stabilized after an additional 10,000 cycles. Fluctuations between high-density (HD) and low-density (LD) phases would be expected near a liquid-liquid phase transition, for small samples, but the high computational cost of these simulations did not allow for exploring this in depth. Overall the results support the view that the MP2 and RPA methods describe the potential energy surface of bulk liquid water accurately and emphasize the value of treating dispersion interactions at the same level as hydrogen-bond interactions.

Figure 3 shows the radial distribution functions (RDF) obtained at the MP2 and RPA level, respectively. The RDFs are in very good agreement with the recent experimental $g_{OO}(r)$ obtained from X-ray diffraction (Skinner et al. 2013). The height of the first peak is overestimated by about 10%, while the features of the long-range part are remarkably well reproduced, being only very slightly more
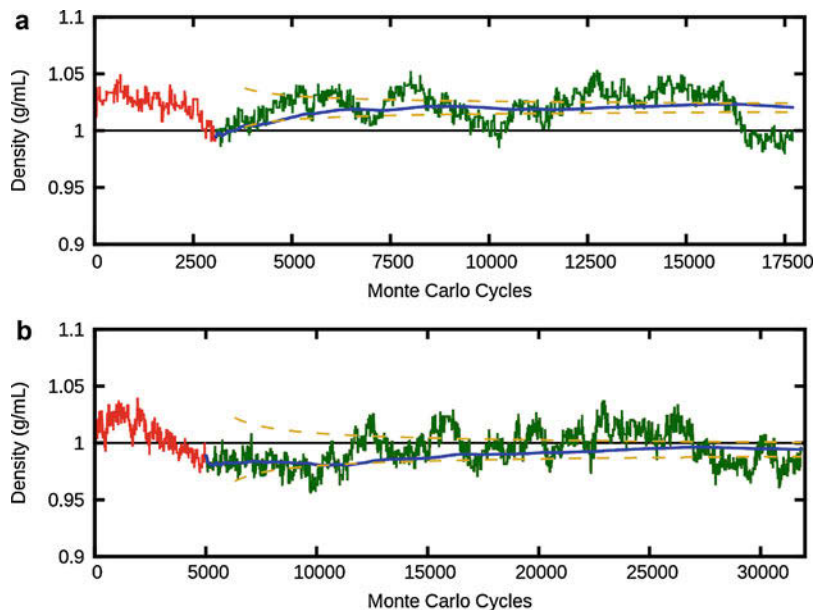
**Fig. 2** The red portion of the plots denotes MC cycles considered for equilibration, the green part refers to that used for the calculation of the average properties, and the blue line shows the running average density. Estimated errors are reported as yellow dashed lines. Panel (**a**) shows results for MP2, panel (**b**) for RPA
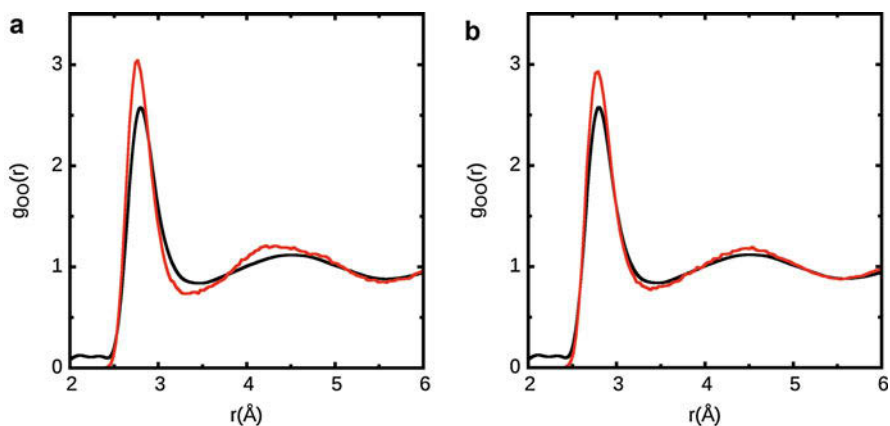


**Fig. 3** Oxygen-oxygen pair radial distribution functions as obtained from the NpT-MC simulations ($T = 295$ K and $p = 1$ bar) at (**a**) MP2 (**b**), RPA level of theory (solid red line). Experimental RDF (solid black line) taken from Skinner et al. (2013)

structured. In order to gain additional information on the local environment around each water molecule in the liquid, the positions of the 4th and 5th neighboring $H_2O$ were analyzed, as they serve as an indicator for the overlap of coordination shells. GGAs and hybrid functionals not corrected for van der Waals interactions give highly structured and significantly less dense (around 10–20%) liquids. The introduction of dispersion improves the liquid description but in general gives too dense water (5–10%) (Del Ben et al. 2015b). In particular the distributions of the 4th and 5th neighboring waters show a clear separation between the first and second coordination shells for GGA functionals, whereas they move closer when vdW interactions are included, and a similar overlapping pattern is found for MP2 and RPA methods.

The accurate prediction of the water density and structure requires thus a balance between vdW and hydrogen-bond interactions. On this basis, it can be argued that the accurate prediction of the water density and structure by RPA and MP2 can be attributed to the fact that these theories provide, for this system, a correct balance between hydrogen-bond and van der Waals interactions. Reaching this balance might be facilitated by the fact that both interactions are obtained from the same level of theory. Note that in general, MP2 and RPA tend to overbind and underbind, respectively, noncovalent complexes as, e.g., found in the S22 set (Jurecka et al. 2006), which covers a broader range of systems, including compounds that are aromatic or have double bonds (Eshuis and Furche 2012; Ren et al. 2012). While these effects are not pronounced for water, it might nevertheless lead to the slightly higher and lower water densities calculated at the MP2 and RPA level, respectively.

To put these results in context, Table 2 summarizes the density and structural features of the RDF of liquid water obtained from NpT-MC simulations at different levels of theory but with consistent choice of simulation parameters (Del Ben et al. 2015b). In addition to functionals including orbital correlation (RPA, MP2), standard GGA and hybrid functionals (PBE, PBE0) with D3 and nonlocal vdW corrections are shown. The reported average values show the slight overestimation

**Table 2** Average density and structural data obtained from the NpT-MC simulations ($T = 295$ K and $p = 1$ bar). The method labeled with optB88-vdW (Klimeš et al. 2010, 2011) represents a functional of the nonlocal van der Waals type. The label D3 stands for a dispersion correction according to Grimme et al. (2010). For the calculated average densities, an error estimation is reported ($\Delta_\rho$)

|            | Density [g/mL] | | 1st Max | | 1st Min | | 2nd Max | |
|------------|------|-------------|-------|----------|-------|----------|-------|----------|
|            | $\rho$ | $\Delta_\rho$ | $r$ [Å] | $g_{OO}(r)$ | $r$ [Å] | $g_{OO}(r)$ | $r$ [Å] | $g_{OO}(r)$ |
| PBE-D3     | 1.055 | 0.006 | 2.73 | 3.07 | 3.25 | 0.69 | 4.43 | 1.21 |
| PBE0-D3    | 1.053 | 0.005 | 2.74 | 2.88 | 3.29 | 0.79 | 4.32 | 1.21 |
| optB88-vdW | 1.081 | 0.003 | 2.74 | 2.94 | 3.34 | 0.80 | 4.31 | 1.21 |
| MP2        | 1.020 | 0.004 | 2.76 | 3.05 | 3.32 | 0.72 | 4.41 | 1.21 |
| RPA        | 0.994 | 0.006 | 2.78 | 2.93 | 3.41 | 0.78 | 4.49 | 1.19 |
| exp.       | 1.00  |       | 2.80 | 2.57 | 3.45 | 0.84 | 4.5  | 1.12 |

when GGA and hybrid functionals with the D3 vdW corrections are used. For the considered nonlocal van der Waals functional, the obtained results show a larger density of ∼8%, and the structure of the liquid is rather well reproduced but with all features shifted to shorter distances, in agreement with previous NVT simulations (Wang et al. 2011). Even though at the limit of statistical accuracy, both the inclusion of the D3 correction and the use of hybrid functionals reduce the height of the first peak of the $g_{OO}(r)$ for PBE. This combined effect of dispersion and exchange has been emphasized recently (DiStasio et al. 2014).

GGA functionals, which underestimate the bandgap, lead to a too polarizable solvent, which in turn should lead to too strong hydrogen bonds. This is similar to the effect of charge transfer from anions to the solvent that has been observed to be too strong for GGAs as compared to hybrids (VandeVondele et al. 2012). The effect of hybrid functionals on the dielectric constant of water ice has been quantified (Schönherr et al. 2014) showing that GGA functionals lead to more polar structures while also overestimating the polarization of these structures. Whereas the effect of bandgap underestimation is already noticeable in neat water, it becomes even more visible in the context of aqueous electrochemistry. Redox levels of various species can be significantly influenced, if water band edges are incorrect; in particular they will be pushed up if these are close to the valence band of the liquid (Adriaanse et al. 2012). Results obtained with RPA and double-hybrid functionals for redox levels near the band edges appear promising (Cheng and VandeVondele 2016). This highlights the importance of going beyond the neat liquid in assessing the performance of electronic structure theory for liquid water.

Finally the equilibrium parameters of ice Ih, calculated at various levels of theory, are reported in Table 3 for comparison (Del Ben et al. 2015b). The results reported are obtained at 0 K neglecting nuclear quantum effects and zero-point energies. Despite being important in many situations, these effects have been shown to have a lesser influence ($\leq$1%) for the equilibrium volume of ice Ih both theoretically (Santra et al. 2013; Pamuk et al. 2012) and experimentally (Röttger et al. 2012). Analogously to the liquid water case, the calculated ice densities are ∼2% larger and ∼1% smaller than experimental results for MP2 and RPA, respectively. At both the MP2 and RPA level, these results show the nontrivial prediction that ice floats on water, with a quantitatively correct ratio of liquid and solid density.

**Table 3** Equilibrium volumes and energies (at 0 K) for ice Ih expressed per molecule without corrections for the quantum nature of the nuclei and zero-point energies. Experimental values are from Hobbs (1974) and Whalley (1984)

|       | $E_{\mathrm{coh}}$ [kJ/mol] | $V_{\mathrm{mol}}$ [Å$^3$] | $\rho$ [g/mL] |
|-------|------------------|------------------|------------|
| PBE   | −62.8            | 30.69            | 0.975      |
| MP2   | −58.7            | 31.34            | 0.955      |
| RPA   | −52.5            | 32.37            | 0.924      |
| Exp.  | −58.9            | 32.05            | 0.933      |

# 4 Outlook

With the methods described here, nonlocal wave function correlation methods can be applied to liquids, solutions, crystals, and simple interfaces. As the available computational resources continue to grow, such applications will become increasingly more routine. At the same time, the algorithms will be further refined and improved. Reduced scaling algorithms (Wilhelm et al. 2016) will make it possible to tackle larger systems. Multiple time-step algorithms are developed that make ab initio molecular dynamics simulations using high-level methods more efficient. These methods are also combined with path integral techniques (Kapil et al. 2016) and bring MP2 and RPA based simulations including nuclear quantum effects within reach.

# References

Adriaanse C, Cheng J, Chau V, Sulpizi M, VandeVondele J, Sprik M (2012) Aqueous redox chemistry and the electronic band structure of liquid water. J Phys Chem Lett 3:3411–3415

Borštnik U, VandeVondele J, Weber V, Hutter J (2014) Sparse matrix multiplication: the distributed block-compressed sparse row library. Parallel Comput 40:47–58

Burow AM, Bates JE, Furche F, Eshuis H (2014) Analytical first-order molecular properties and forces within the adiabatic connection random phase approximation. J Chem Theory Comput 10:180–194

Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. Phys Rev Lett 55:2471–2474

Cheng J, VandeVondele J (2016) Calculation of electrochemical energy levels in water using the random phase approximation and a double hybrid functional. Phys Rev Lett 116:086402

Del Ben M, Hutter J, VandeVondele J (2012) Second-order Møller–Plesset perturbation theory in the condensed phase: an efficient and massively parallel Gaussian and plane waves approach. J Chem Theory Comput 8:4177–4188

Del Ben M, Hutter J, VandeVondele J (2013a) Electron correlation in the condensed phase from a resolution of identity approach based on the Gaussian and plane waves scheme. J Chem Theory Comput 9:2654–2671

Del Ben M, Schönherr M, Hutter J, VandeVondele J (2013b) Bulk liquid water at ambient temperature and pressure from MP2 theory. J Phys Chem Lett 4:3753–3759

Del Ben M, Hutter J, VandeVondele J (2015a) Forces and stress in second order Møller-Plesset perturbation theory for condensed phase systems within the resolution-of-identity Gaussian and plane waves approach. J Chem Phys 143:102803

Del Ben M, Hutter J, VandeVondele J (2015b) Probing the structural and dynamical properties of liquid water with models including non-local electron correlation. J Chem Phys 143:054506

Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density functional for general geometries. Phys Rev Lett 92:246401

DiStasio RA, Santra B, Li Z, Wu X, Car R (2014) The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. J Chem Phys 141:084502

Dunlap BI, Connolly JWD, Sabin JR (1979) On some approximations in applications of X alpha theory. J Chem Phys 71:3396–3402

Dunning TH (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J Chem Phys 90:1007–1023

Errington JR, Debenedetti PG (2001) Relationship between structural order and the anomalies of liquid water. Nature 409:318–321

Eshuis H, Furche F (2012) Basis set convergence of molecular correlation energy differences within the random phase approximation. J Chem Phys 136:084105

Eshuis H, Yarkony J, Furche F (2010) Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. J Chem Phys 132:234114

Eshuis H, Bates J, Furche F (2012) Electron correlation methods based on the random phase approximation. Theor Chem Acc 131:1084

Feyereisen M, Fitzgerald G, Komornicki A (1993) Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. Chem Phys Lett 208:359–363

Furche F (2008) Developing the random phase approximation into a practical post-Kohn–Sham correlation model. J Chem Phys 129:114105

García-González P, Fernández JJ, Marini A, Rubio A (2007) Advanced correlation functionals: application to bulk materials and localized systems. J Phys Chem A 111:12458–12465

Goedecker S, Teter M, Hutter J (1996) Separable dual-space Gaussian pseudopotentials. Phys Rev B 54:1703–1710

Goerigk L, Grimme S (2011) Efficient and accurate double-hybrid-meta-GGA density functionals: evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. J Chem Theory Comput 7:291–309

Grimme S (2006) Semiempirical hybrid density functional with perturbative second-order correlation. J Chem Phys 124:034108

Grimme S, Steinmetz M (2016) A computationally efficient double hybrid density functional based on the random phase approximation. Phys Chem Chem Phys 18:20926–20937

Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104

Guidon M, Schiffmann F, Hutter J, VandeVondele J (2008) Ab initio molecular dynamics using hybrid density functionals. J Chem Phys 128:214104

Guidon M, Hutter J, VandeVondele J (2009) Robust periodic Hartree–Fock exchange for large-scale simulations using Gaussian basis sets. J Chem Theory Comput 5:3010–3021

Handy NC, Schaefer HF (1984) On the evaluation of analytic energy derivatives for correlated wave functions. J Chem Phys 81:5031–5033

Hobbs P (1974) Ice physics. Clarendon Press, Oxford

Hutter J (2012) Car-Parrinello molecular dynamics. WIREs Comput Mol Sci 2:604–612

Iftimie R, Salahub D, Wei D, Schofield J (2000) Using a classical potential as an efficient importance function for sampling from an ab initio potential. J Chem Phys 113:4852–4862

Jurecka P, Sponer J, Cerny J, Hobza P (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. Phys Chem Chem Phys 8:1985–1993

Kapil V, VandeVondele J, Ceriotti M (2016) Accurate molecular dynamics and nuclear quantum effects at low cost by multiple steps in real and imaginary time: using density functional theory to accelerate wavefunction methods. J Chem Phys 144:054111

Klimeš J, Bowler D, Michaelides A (2010) Chemical accuracy for the van der Waals density functional. J Phys Condens Matter 22:022201

Klimeš J, Bowler D, Michaelides A (2011) Van der Waals density functionals applied to solids. Phys Rev B 83:1–13

Laasonen K, Sprik M, Parrinello M, Car R (1993) "Ab initio" liquid water. J Chem Phys 99:9080–9089

Lippert G, Hutter J, Parrinello M (1997) A hybrid Gaussian and plane wave density functional scheme. Mol Phys 92:477–488

Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Leiden

McGrath MJ, Siepmann JI, Kuo IFW, Mundy CJ, VandeVondele J, Sprik M, Hutter J, Mohamed F, Krack M, Parrinello M (2005a) Toward a Monte Carlo program for simulating vapor-liquid phase equilibria from first principles. Comput Phys Commun 169:289–294

McGrath MJ, Siepmann JI, Kuo IFW, Mundy CJ, VandeVondele J, Hutter J, Mohamed F, Krack M (2005b) Isobaric-isothermal Monte Carlo simulations from first principles: application to liquid water at ambient conditions. ChemPhysChem 6:1894–1901

Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46:618–622

Neese F, Schwabe T, Grimme S (2007) Analytic derivatives for perturbatively corrected "double hybrid" density functionals: theory, implementation, and applications. J Chem Phys 126:124115

Nielsen OH, Martin RM (1985) Quantum-mechanical theory of stress and force. Phys Rev B 32:3780–3791

Nilsson A, Pettersson L (2011) Perspective on the structure of liquid water. Chem Phys 389:1–34

Paier J, Ren X, Rinke P, Scuseria GE, Grüneis A, Kresse G, Scheffler M (2012) Assessment of correlation energies based on the random-phase approximation. New J Phys 14:043002

Pamuk B, Soler JM, Ramírez R, Herrero CP, Stephens PW, Allen PB, Fernández-Serra MV (2012) Anomalous nuclear quantum effects in ice. Phys Rev Lett 108:193003

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Perdew JP, Ruzsinszky A, Tao J, Staroverov VN, Scuseria GE, Csonka GI (2005) Prescription for the design and selection of density functional approximations: more constraint satisfaction with fewer fits. J Chem Phys 123:062201

Rahman A, Stillinger FH (1971) Molecular dynamics study of liquid water. J Chem Phys 55:3336–3359

Ramberger B, Schäfer T, Kresse G (2017) Analytic interatomic forces in the random phase approximation. Phys Rev Lett 118:106403

Ren X, Rinke P, Joas C, Scheffler M (2012) Random-phase approximation and its applications in computational chemistry and materials science. J Mater Sci 47:7447

Röttger K, Endriss A, Ihringer J, Doyle S, Kuhs WF (2012) Lattice constants and thermal expansion of $H_2O$ and $D_2O$ Ice I$h$ between 10 and 265K. Addendum. Acta Crystallogr Sec B 68:91

Rybkin VV, VandeVondele J (2016) Spin-unrestricted second-order Møller–Plesset (MP2) forces for the condensed phase: from molecular radicals to f-centers in solids. J Chem Theory Comput 12:2214–2223

Santra B, Klimeš J, Tkatchenko A, Alfè D, Slater B, Michaelides A, Car R, Scheffler M (2013) On the accuracy of van der Waals inclusive density-functional theory exchange-correlation functionals for ice at ambient and high pressures. J Chem Phys 139:154702

Schmidt J, VandeVondele J, Kuo IFW, Sebastiani D, Siepmann JI, Hutter J, Mundy CJ (2009) Isobaric–isothermal molecular dynamics simulations utilizing density functional theory: an assessment of the structure and density of water at near–ambient conditions. J Phys Chem B 113:11959–11964

Schönherr M, Slater B, Hutter J, VandeVondele J (2014) Dielectric properties of water ice, the ice Ih/XI phase transition, and an assessment of density functional theory. J Phys Chem B 118:590–596

Scuseria GE, Henderson TM, Sorensen DC (2008) The ground state correlation energy of the random phase approximation from a ring coupled cluster doubles approach. J Chem Phys 129:231101

Skinner LB, Huang C, Schlesinger D, Pettersson LGM, Nilsson A, Benmore CJ (2013) Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. J Chem Phys 138:074506

Soper AK (2013) The radial distribution functions of water as derived from radiation total scattering experiments: is there anything we can say for sure? ISRN Phys Chem 2013:279463

Todorova T, Seitsonen AP, Hutter J, Kuo IFW, Mundy CJ (2006) Molecular dynamics simulation of liquid water: hybrid density functionals. J Phys Chem B 110:3685–3691

Vahtras O, Almlöf J, Feyereisen M (1993) Integral approximations for LCAO-SCF calculations. Chem Phys Lett 213:514–518

VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J (2005a) Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. Comput Phys Commun 167:103–128

VandeVondele J, Mohamed F, Krack M, Hutter J, Sprik M, Parrinello M (2005b) The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. J Chem Phys 122:014515

VandeVondele J, Troester P, Tavan P, Mathias G (2012) Vibrational spectra of phosphate ions in aqueous solution probed by first-principles molecular dynamics. J Phys Chem A 116: 2466–2474

Wang J, Román-Pérez G, Soler JM, Artacho E, Fernández-Serra MV (2011) Density, structure, and dynamics of water: the effect of van der Waals interactions. J Chem Phys 134:024516

Weigend F, Häser M (1997) RI-MP2: first derivatives and global consistency. Theor Chem Acc 97:331–340

Weigend F, Häser M, Patzelt H, Ahlrichs R (1998) RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. Chem Phys Lett 294:143–152

Whalley E (1984) Energies of the phases of ice at zero temperature and pressure. J Chem Phys 81:4087–4092

Whitten JL (1973) Coulombic potential energy integrals and approximations. J Chem Phys 58:4496–4501

Wilhelm J, Seewald P, Del Ben M, Hutter J (2016) Large-scale cubic-scaling random phase approximation correlation energy calculations using a Gaussian basis. J Chem Theory Comput 12:5851–5859

# Accelerated Molecular Dynamics for Ab Initio Electronic Simulations

# 26

Guglielmo Mazzola and Sandro Sorella

## Contents

**Abstract**

A recently proposed method for accelerating current molecular dynamics algorithms, used for the simulation of classical particles at finite temperatures, is reviewed (Mazzola and Sorella, Phys Rev Lett 118:015703, 2017). This method is based on an efficient implementation of a first-order Langevin dynamics modified in a way to reduce the autocorrelation times and the time step error for the integration of the stochastic equations of motion. This work represents an improvement upon previously known algorithms that, on one hand, are too much simplified to be used in realistic simulations and, on the other hand, are too much complicated and computationally demanding for their practical implementations.

G. Mazzola
Theoretische Physik, ETH Zurich, Zurich, Switzerland
e-mail: gmazzola@phys.ethz.ch

S. Sorella (✉)
International School for Advanced Studies (SISSA), Trieste, Italy
e-mail: sorella@sissa.it

The details of the method are presented with few applications to standard test cases on Lennard-Jones models at various temperatures. In particular it is shown that this technique represents an ideal tool for ab initio molecular dynamics, when the Born-Oppenheimer energy surface is estimated by computationally demanding methods, such as, for instance, the quantum Monte Carlo stochastic approach.

# 1    Introduction

One of the simplest, but nevertheless robust, methods for simulating a given number $N$ of classical particles in a finite volume $V$ for a temperature $T$, namely, for computing physical quantities in the so-called $NVT$ ensemble (Allen and Tildesley 1987; Tuckerman 2010) is given by the *first-order* Langevin dynamics, which is defined by the following set of stochastic differential equations:

$$\frac{d\mathbf{R}(t)}{dt} = \mathbf{f}[\mathbf{R}(t)] + \boldsymbol{\eta}(t), \tag{1}$$

where $\mathbf{R}(t)$ is a (time dependent) $p$ dimensional vector with components $R_j$ ($j = 1, \ldots, p$, with $p = N \times d$, $d$ being the spatial dimensionality). The classical potential $V(\mathbf{R})$ defines the deterministic "force" with components $f_j(\mathbf{R})$:

$$f_j(\mathbf{R}) = -\frac{\partial V(\mathbf{R})}{\partial R_j}; \tag{2}$$

whereas $\boldsymbol{\eta}(t)$ is a random vector with components $\eta_j(t)$ ($j = 1, \ldots, p$), with vanishing mean value and no correlations between components $i \neq j$ and times $t \neq t'$:

$$\langle \eta_j(t) \rangle = 0, \tag{3}$$

$$\langle \eta_i(t)\eta_j(t') \rangle = 2T\delta_{i,j}\delta(t - t'), \tag{4}$$

where $\langle \ldots \rangle$ indicates the expectation value. This conventional first-order Langevin dynamics (CFOLD) (1) is substantially different from the Newton equations of motions that are *second-order* (deterministic) differential equations connecting the actual force to the acceleration of the particles. The advantage of using first-order equations will be clear in the following because of their apparent simplicity. On one hand, they are analytically less involved, and, on the other hand, they contain less parameters, namely, the mass of the particles and their damping coefficients. The latter quantities have to be introduced in the Newton-like equations in order to sample correctly the NVT ensemble, because they provide a simple solution to the well-known ergodicity problems of weakly interacting harmonic systems and, in this way, improve the efficiency of the sampling (Ceriotti et al. 2009; Kühne et al. 2009).

Contrary to ordinary differential equations, the stochastic differential equations do not provide a unique solution starting from a given initial condition, but the possible solutions acquire a statistical meaning: starting at $t = t_0$ from $\mathbf{R}_0 \equiv \mathbf{R}(t_0)$, several stochastic trajectories are possible depending on the particular realization of the noise. Thus, we are naturally led to define the probability $P(\mathbf{R}, t)$ to find a given configuration $\mathbf{R}$ at time $t$, with the initial condition:

$$P(\mathbf{R}, t_0) = \delta(\mathbf{R} - \mathbf{R}_0). \tag{5}$$

After an equilibration time, $P(\mathbf{R}, t)$ converges to the equilibrium distribution $P_{\text{eq}}(\mathbf{R})$ that is independent of the initial condition $\mathbf{R}_0$. Then, the solution $\mathbf{R}_n$ at discrete times $t_n$ of the differential equations (1) provides a large number $M$ of samples that can be used to compute statistically any correlation function $O(\mathbf{R})$:

$$\int \mathbf{dR} \, O(\mathbf{R}) \, P_{\text{eq}}(\mathbf{R}) \approx \frac{1}{M} \sum_n O(\mathbf{R}_n) \tag{6}$$

similarly to the ordinary Monte Carlo method.

The main purpose of this chapter is to show that, by using the simpler formalism of the first-order stochastic equation, it is possible to solve exactly, within the harmonic approximation, the problem of long autocorrelation times, usually affecting the implementation of molecular dynamics in complex systems such as water at ambient conditions or complex processes such as protein folding. The latter phenomenon takes place within the time scale of $\mu s$, much larger than the period of the molecular vibrations of the order of the $fs$ (Scheraga et al. 2007). To this end, we have to introduce a more general approach, already known in literature, starting with the covariant version of the first-order Langevin dynamics equations (LDE):

$$\dot{\mathbf{R}} = S^{-1}[\mathbf{R}(t)]\mathbf{f}[\mathbf{R}(t)] + \boldsymbol{\eta}$$
$$\langle \eta_i \eta_j \rangle = 2T\delta(t - t')S_{ij}^{-1}[\mathbf{R}(t)] \tag{7}$$

where $S[\mathbf{R}(t)]$ is, for the time being, an arbitrary positive-definite symmetric matrix – i.e., with all eigenvalues strictly positive – that is used to accelerate the dynamics in the way we will discuss in the following sections. Notice also that this matrix is explicitly dependent on the atomic positions, a peculiar but very important property of this approach; otherwise, this method coincides with the standard accelerated dynamics introduced long time ago by Parisi (1984). Moreover, for $S_{ij} = \delta_{ij}$, Eq. (7) describes the conventional first-order Langevin dynamics discussed at the beginning.

The equilibrium probability implied by these stochastic differential equations, containing a position dependent matrix $S[\mathbf{R}(t)]$, is given by a generalized Boltzmann distribution:

$$P_{\text{eq}}(\mathbf{R}) = \frac{1}{\mathscr{Z}} |S[\mathbf{R}(t)]|^\gamma \exp\left[-\frac{V(\mathbf{R})}{T}\right],\tag{8}$$

where $|S[\mathbf{R}(t)]|$ is the determinant of the matrix $S[\mathbf{R}(t)]$ and $\mathscr{Z}$ is the partition function, needed for the normalization condition of the probability:

$$\mathscr{Z} = \int \mathbf{dR} |S[\mathbf{R}(t)]|^\gamma \exp\left[-\frac{V(\mathbf{R})}{T}\right].\tag{9}$$

In principle the LDE of Eq. (7) does not depend on the parameter $\gamma$, while the equilibrium distribution is explicitly dependent on it. At first sight, this may appear in contradiction, but, as we will see in the following, this dependency may arise due to the arbitrariness of the continuous limit of stochastic differential equations, as there are several consistent ways to integrate Eq. (7). Notice that for $\gamma = 0$, we obtain the conventional canonical Boltzmann distribution that can be used for our purposes. Another important case is when $\gamma = 1/2$, the matrix $S$ representing the metric of the space, and the LDE acquire the covariant property, namely, they are independent of any nonlinear transformation $\mathbf{R} \to \mathbf{R}'$ of the variables in the non-euclidean space defined by the metric $S[\mathbf{R}(t)]$, namely, the infinitesimal distance $ds$ between two points of the space is given by $ds^2 = \sum_{ij} S_{i,j}[\mathbf{R}] dR_i dR_j$. The motivation for introducing an acceleration matrix $S[\mathbf{R}(t)]$ in the equations becomes clear if we consider the limiting $T = 0$ case, where the finite temperature Eq. (7) reduces to a structural relaxation equation. If we first discretize Eq. (7), with a finite time step $\Delta$, and consider an harmonic problem defined by a potential

$$V_H(\mathbf{R}) = \frac{1}{2} \sum_{i,j} H_{i,j} (\mathbf{R} - \mathbf{R}^{eq})_i (\mathbf{R} - \mathbf{R}^{eq})_j \tag{10}$$

where $H$ is the Hessian matrix and $\mathbf{R}^{eq}$ the equilibrium positions, we obtain that

$$\mathbf{R}(t + \Delta) = \mathbf{R}(t) + \Delta S^{-1} \mathbf{f}(t).\tag{11}$$

Thus, if $S[\mathbf{R}] = H$ and $\Delta = 1$, it is possible to reach the minimum $\mathbf{R}_{eq}$, starting from any arbitrary configuration $\mathbf{R}(t)$, in only one step, simply because $\mathbf{S}^{-1}\mathbf{f}(t) = -(\mathbf{R}(t) - \mathbf{R}^{eq})$ in this case. On the contrary, the standard steepest-descent algorithm, with $S[\mathbf{R}]$ being the identity matrix, is affected by a slowing down controlled by the condition number $K_{\text{cond}} = K_{\max}/K_{\min}$ of the matrix $H$, i.e., the ratio between its maximum ($K_{\max}$) and minimum non-zero ($K_{\min}$) eigenvalues.

In the following we generalize this approach from structural optimization to the finite temperature dynamics of our interest. The main difficulty consists in finding an efficient iterative scheme able to produce the unbiased sampling while retaining the property of decreasing the autocorrelation time by means of a suitable position dependent matrix $S[\mathbf{R}]$. In the next section, we develop such scheme in a general way, without specifying the matrix $S[\mathbf{R}]$ but keeping in mind that, for practical

calculations, this matrix could be a positive-definite matrix that at low temperatures is very close to the Hessian for optimal performances. This approach is similar in spirit to the mass tensor scheme (Bennett 1975; Tsuchida 2015) where a mass tensor **M** plays the role of our acceleration matrix $S$. However the method that we present here becomes much simpler when it is necessary to consider acceleration matrices ($M$ or $S$) that are explicitly dependent on **R**, because the algorithm we propose does not contain any cumbersome differentiation of $S$. On the other hand, recent attempts based on damped molecular dynamics (Ceriotti et al. 2010) cannot have optimal efficiency in the harmonic case but can reduce the correlation time only by a factor proportional to $\sqrt{K_{\text{cond}}}$ (Tassone et al. 1994). They should be therefore much less efficient (by a factor proportional to $\sqrt{K_{\text{cond}}}$) than the algorithm we discuss here.

## 2    Fokker-Planck Equation and Approach to the Canonical Distribution

Quite generally, conventional differential equations are integrated approximately at discrete short time intervals $\Delta$, by introducing a so-called time step error that is unavoidable in any realistic simulation. Unfortunately stochastic differential equations like the LDE in Eq. (7) are not well defined because the noise introduces nonanalytic terms proportional to $\sqrt{\Delta}$, yielding discretized equations depending upon arbitrary choices (e.g., Ito or Stratonovich). In our opinion the correct way to deal with stochastic differential equation is to start just from a given, though arbitrarily chosen, discretized version of the equation and show that the associated probability distribution $P(\mathbf{R}, t)$ satisfies a well-defined Fokker-Planck equation, equilibrating to the canonical distribution $P_{\text{eq}}(\mathbf{R})$ in the long time limit.

We start therefore from the integration of Eq. (7), at discrete time $t = n\Delta$, that is conventionally adopted in textbooks (Risken 1996) and is given by the following rather involved expression

$$\mathbf{R}(t + \Delta)_j = \mathbf{R}(t)_j + \Delta \left( S^{-1}(\mathbf{R})\mathbf{f_R} \right)_j + T\Delta \left( \sum_i \partial_i S_{ji}^{-1}(\mathbf{R}) \right) + \sqrt{2T\Delta} z_j(t)$$

$$\langle z_i(t)z_j(t) \rangle = S_{i,j}^{-1}(\mathbf{R}(t)) \tag{12}$$

$$(\mathbf{f_R})_j = -\partial_j V(\mathbf{R}) + \gamma T \partial_j \ln |S(\mathbf{R})| \tag{13}$$

where, as it will be shown in the following, this discretization provides an equilibrium distribution for $\Delta \to 0$ of the form given in Eq. (8). In the following we are interested to the canonical distribution, defined in the standard Euclidean metric, and therefore we have to keep in mind that our interest is for $\gamma = 0$, though all the forthcoming derivation remains valid for any value of $\gamma$.

The above Markov chain univocally defines a discretized master equation for the probability function $P(\mathbf{R}, t)$ that, in the limit $\Delta \to 0$, becomes a Fokker-Planck one of the following form:

$$\partial_t P(\mathbf{R}, t) = \sum_j \partial_j \left\{ -\left[ S^{-1}(\mathbf{R}) f_{\mathbf{R}} \right]_j P(\mathbf{R}, t) + T \sum_i \left[ S_{j,i}^{-1}(\mathbf{R}) \partial_i \right] P(\mathbf{R}, t) \right\}$$

$$(14)$$

In order to find the equilibrium distribution, it is enough to equate to zero the term between braces which immediately gives $P(\mathbf{R}, t)$ up to a constant, in turn determined by the normalization condition of probabilities, yielding:

$$P_{eq}(\mathbf{R}, t) = \frac{|S(\mathbf{R})|^\gamma \exp(-V(\mathbf{R})/T)}{\mathscr{Z}},$$

$$(15)$$

namely, the desired distribution. Unfortunately the Markov chain of Eq. (12) is not practical, because it contains, what in the following will be named the "cumbersome term":

$$\Gamma_j(\mathbf{R}) = \sum_i \partial_i S_{j,i}^{-1}(\mathbf{R})$$

$$(16)$$

Indeed, this term is computationally very demanding, because the calculation of the inverse of a matrix takes the order of $p^3$ operations, as well as each derivative of $S^{-1}$ with respect to any variable $i$, e.g., by using a finite difference method. Thus, in order to make the summation overall $i$ for each $j$ in the above equation, we end up with an algorithm scaling in most cases as the fourth power of $p$, unless for particularly simple cases. Moreover the expressions for the inverse derivatives become so much complicated that they are very difficult to implement in practice (Mazzola et al. 2012), especially within ab initio approaches such as quantum Monte Carlo (Foulkes et al. 2001), DFT (Kohn and Sham 1965), or quantum chemistry wave function methods (Dykstra et al. 2005).

## 3 Avoiding the "Cumbersome Term": A First Simplified Algorithm

Before deriving the final convenient expression for sampling in the most efficient way the canonical distribution by the proposed accelerated Langevin dynamics, we consider the following Markov chain that does not require the evaluation of the "cumbersome term" $\Gamma_j(\mathbf{R})$:

$$\mathbf{y} = \mathbf{R} + \sqrt{2T\Delta}\mathbf{z}(t)$$

$$\mathbf{R}' = \mathbf{y} + \Delta S^{-1}(\mathbf{R})\mathbf{f}_{\mathbf{R}} - \frac{1}{2}S^{-1}(\mathbf{R})\left[ S(\mathbf{y}) - S(\mathbf{R}) \right](\mathbf{y} - \mathbf{R})$$

$$\langle z_i(t) z_j(t) \rangle = S_{i,j}^{-1}(\mathbf{R})$$

$$(17)$$

where here $\mathbf{R} = \mathbf{R}(t)$ and $\mathbf{R}' = \mathbf{R}(t + \Delta)$. We will show in the following that the above Markov chain implies the same Fokker-Planck equation (14) corresponding the much more involved discretization in Eq. (12), and therefore the equilibrium distribution for $\Delta \to 0$ will be the correct one. In this way it is possible to avoid the "cumbersome term," with a minor computational effort, namely, by calculating the matrix $S$ *twice* for each time step.

Let's therefore proceed with the main proof of this section. The Markov chain in Eq. (17) defines in a unique way the conditional probability density of having $\mathbf{R}' = \mathbf{R}(t + \Delta)$ given $\mathbf{R}(t) = \mathbf{R}$:

$$
\begin{aligned}
K(\mathbf{R}'|\mathbf{R}) = \int dz^p \mu_{\mathbf{R}}(\mathbf{z}) \delta \Big\{ &\mathbf{R}' - \mathbf{R} - \Delta S^{-1}(\mathbf{R})\mathbf{f_R} - \sqrt{2T\Delta}\mathbf{z} \\
&+ \frac{1}{2} S^{-1}(\mathbf{R}) \Big[ S(\mathbf{R} + \sqrt{2T\Delta}\mathbf{z}(t)) - S(\mathbf{R}) \Big] \sqrt{2T\Delta}\mathbf{z}(t) \Big\}
\end{aligned}
$$

$$(18)$$

where the normalized Gaussian probability density $\mu_{\mathbf{R}}(z)$ is given, according to Eq. (17), by:

$$
\begin{aligned}
\mu_{\mathbf{R}}(\mathbf{z}) &= \frac{\exp\left[-\frac{1}{2}(z, S(\mathbf{R})z)\right]}{\int\int dz^p \exp\left[-\frac{1}{2}(z, S(\mathbf{R})z)\right]} \\
&= (2\pi)^{-\frac{p}{2}} \exp\left[-\frac{1}{2}(z, S(\mathbf{R})z) + \frac{1}{2}Tr[\ln S(\mathbf{R})]\right]
\end{aligned}
$$

$$(19)$$

where in the latter equation, we have used a well-known property relating the multi-dimensional Gaussian integrals to the determinant $|S(\mathbf{R})|$ written, for convenience, as $\exp Tr[\ln S(\mathbf{R})]$. Here and henceforth we denote by $(a, b)$ the scalar product of two $p-$ dimensional real vectors, and matrix-vector multiplication is understood when a matrix appears before (from left to right) a given vector. We want to obtain a Fokker-Planck equation in the limit of $\Delta \to 0$. To this purpose we write the master equation:

$$
P_{n+1}(\mathbf{R}') = \int dR^p K(\mathbf{R}'|\mathbf{R}) P_n(\mathbf{R}) \tag{20}
$$

and employ the integration in $dR^p$ after substituting the expression of $K(\mathbf{R}'|\mathbf{R})$ given above. In all the forthcoming derivation, all the effort is spent to extract the leading term $O(\Delta)$ of the above master equation that will define in a unique way the Fokker-Planck equation that is in turn necessary to establish the equilibrium distribution. For this purpose we solve the argument of the $\delta$ function, by replacing $\mathbf{R}$ with $\mathbf{R}'$ when it is allowed at the leading order in $\Delta$:

$$\mathbf{R}(\mathbf{R}') = \mathbf{R}' - \sqrt{2T\varDelta}\mathbf{z} - \varDelta S^{-1}(\mathbf{R}')\mathbf{f}_{\mathbf{R}'}$$
$$+ \frac{1}{2}S^{-1}(\mathbf{R}')\left[S(\mathbf{R}' + \sqrt{2T\varDelta}\mathbf{z}) - S(\mathbf{R}')\right]\sqrt{2T\varDelta}\mathbf{z} + o(\varDelta) \qquad (21)$$

we obtain that the master equation for the evolution of the probability is explicitly given:

$$P_{n+1}(\mathbf{R}') = \int dz^p J_\varDelta(\mathbf{R}(\mathbf{R}'))\mu_{\mathbf{R}(\mathbf{R}')}(\mathbf{z})P_n\left(\mathbf{R}(\mathbf{R}')\right) \qquad (22)$$

where $J_\varDelta(\mathbf{R}(\mathbf{R}')) = 1 + \varDelta B(\mathbf{R}') + o(\varDelta)$ is the Jacobian of the transformation of Eq. (21) that can be expanded in $\varDelta$ with a well-defined expression for $B(\mathbf{R}')$ that we do not explicitly write in the following, because, as we will see soon, it is not important for the derivation. $\mu_{\mathbf{R}(\mathbf{R}')}(\mathbf{z})$ comes from the original probability density $\mu_{\mathbf{R}}(\mathbf{z})$ in Eq. (19) once $\mathbf{R}$ is substituted with the $\mathbf{z}$-dependent expression given in Eq. (21). Indeed, by substituting the transformation of Eq. (21) in Eq. (22), and expanding the latter equation to the leading order in $\varDelta$, we obtain the following expression:

$$P_{n+1}\left(\mathbf{R}'\right) = \left[1 + \varDelta C(\mathbf{R}')\right]P_n(\mathbf{R}') + \int dz^p \mu_{\mathbf{R}(\mathbf{R}')}(\mathbf{z})\left\{-\sum_j\left\{\sqrt{2T\varDelta}z_j\right.\right.$$
$$+\varDelta\left[S^{-1}(\mathbf{R}')\left(\mathbf{f}_{\mathbf{R}'} - \frac{S(\mathbf{R}' + \sqrt{2T\varDelta}\mathbf{z}) - S(\mathbf{R}')}{2\varDelta}\sqrt{2T\varDelta}\mathbf{z}\right)\right]_j\right\}\partial_j P_n(\mathbf{R}')$$
$$\left. + \varDelta T\sum_{i,j}z_i z_j \partial_i \partial_j P_n(\mathbf{R}')\right\} + o(\varDelta) \qquad (23)$$

In the above iteration of Eq. (23), there is therefore a term that simply multiplies $P_n(\mathbf{R}')$ by a function:

$$1 + \varDelta C(\mathbf{R}') \qquad (24)$$

where $C(\mathbf{R}')$ is rather involved and comes from the expansion in small $\varDelta$ of all the factors ($J$ and $\mu$) of the integrand in Eq. (22) multiplying $P(\mathbf{R}(\mathbf{R}'))$:

$$C(\mathbf{R}')\varDelta = -1 + \int dz^p J_\varDelta(\mathbf{R}(\mathbf{R}'))\mu_{\mathbf{R}(\mathbf{R}')}(\mathbf{z}) = \varDelta B(\mathbf{R}') - 1 + \int dz^p \mu_{\mathbf{R}(\mathbf{R}')}(\mathbf{z}) + o(\varDelta) \qquad (25)$$

Notice that $\mathbf{R}(\mathbf{R}')$ depends on the random variable $\mathbf{z}$ via Eq. (21), and therefore the term $\int dz^p \mu_{\mathbf{R}(\mathbf{R}')}(\mathbf{z})$ in the above equation is nontrivial and different from one by $O(\varDelta)$. We will not attempt to calculate this term, as well as $B(\mathbf{R}')$, but derive it from the conservation of the normalization condition of the probability.

As evident in Eq. (23), there are three different contributions depending on the order of the derivatives. The term that couples to the first derivative of $P_n(\mathbf{R}')$ reads:

$$-\Delta \sum_j \int dz^P \mu_{\mathbf{R}'}(\mathbf{z})$$

$$\left[ S^{-1}(\mathbf{R}') \left( \mathbf{f}_{\mathbf{R}'} - \frac{S(\mathbf{R}' + \sqrt{2T\Delta}\mathbf{z}) - S(\mathbf{R}')}{2\Delta} \sqrt{2T\Delta}\mathbf{z} \right) \right]_j \partial_j P_n(\mathbf{R}') \qquad (26)$$

$$-\int dz^P \mu_{\mathbf{R}'}(\mathbf{z}) \sum_{i,j,k,l} (2\Delta T)z_j$$

$$\left[ \frac{1}{2} z_k \left( \partial_i S_{k,l}(\mathbf{R}') \right) z_l z_i - \frac{1}{2} S_{k,l}^{-1}(\mathbf{R}') \left( \partial_i S_{kl}(\mathbf{R}') \right) z_i \right] \partial_j P_n(\mathbf{R}') \qquad (27)$$

where the latter equation comes from the expansion of $\mu(\mathbf{R}(\mathbf{R}'))$ at leading order $O(\sqrt{\Delta})$, where, to this end, we have also used the following relation:

$$Tr[\ln S\left(\mathbf{R}(\mathbf{R}')\right)] = Tr[\ln S(\mathbf{R}')] - \sum_i Tr[S^{-1}(\mathbf{R}')\partial_i S(\mathbf{R}')]\sqrt{2T\Delta}z_i + O(\Delta)$$

$$= Tr[\ln S(\mathbf{R}')] - \sqrt{2T\Delta} \sum_{i,k,l} S_{kl}^{-1}(\mathbf{R}')\partial_i S_{kl}(\mathbf{R}')z_i + O(\Delta) \quad (28)$$

and in the last equation, we have used that $S^{-1}$ is symmetric because $S$ is symmetric. By carrying out the simple integration in $dz^P$, i.e., by replacing $\int dz^P \mu_{\mathbf{R}'}(\mathbf{z})z_j =<z_j> = 0$ and $\int dz^P \mu_{\mathbf{R}'}(\mathbf{z})z_i z_j = <z_i z_j> = S_{i,j}^{-1}(\mathbf{R}')$ and by applying the Wick's theorem for the integration of the higher-order polynomial involved, i.e., $<z_j z_k z_l z_i> = <z_j z_k><z_l z_i> + <z_j z_l><z_k z_i> + <z_j z_i><z_k z_l>$, we obtain that Eq. (27) reads:

$$-\Delta T \sum_{i,j,k,l} \left[ \left( S_{jk}^{-1}S_{li}^{-1} + S_{jl}^{-1}S_{ki}^{-1} + S_{ji}^{-1}S_{kl}^{-1} - S_{kl}^{-1}S_{ji}^{-1} \right) \partial_i S_{kl} \right] (\mathbf{R}')\partial_j P_n(\mathbf{R}')$$

$$= 2\Delta T \sum_{i,j} \left( \partial_i S_{j,i}^{-1}(\mathbf{R}') \right) \partial_j P_n(\mathbf{R}')$$

$$= 2T\Delta \sum_j \Gamma_j(\mathbf{R})\partial_j P(\mathbf{R}') \qquad (29)$$

where, by $S + \delta S = S(I + S^{-1}\delta S) \rightarrow (S + \delta S)^{-1} = S^{-1} - S^{-1}\delta S S^{-1} + o(\delta S) \rightarrow \partial_i S_{ji}^{-1} = -\left[ S^{-1}(\partial_i S)S^{-1} \right]_{ji}$, we easily verify that the LHS and RHS of the above Eq. (29) are consistent, as, for instance, by using that $S$ is a symmetric matrix, we have that:

$$\sum_{kl} S_{jl}^{-1}(\mathbf{R}') S_{ki}^{-1}(\mathbf{R}') \partial_i S_{kl}(\mathbf{R}') = \sum_{kl} S_{jl}^{-1}(\mathbf{R}') S_{ki}^{-1}(\mathbf{R}') \partial_i S_{lk}(\mathbf{R}') = -\partial_i S_{ji}^{-1}(\mathbf{R}') \tag{30}$$

Thus this term partially cancels with the contribution coming from the expansion in small $\Delta$ of the term

$$\left[ S(\mathbf{R}' + \sqrt{2T\Delta}\mathbf{z}) - S(\mathbf{R}') \right]_{k,l} = \sqrt{2T\Delta} \sum_i \partial_i S_{kl}(\mathbf{R}') z_i \tag{31}$$

present in Eq. (26). Indeed, in the Fokker-Planck equation, the term proportional to $\partial_j P$ coming from the Eq. (26) acquires a contribution:

$$T\Delta \int dz^p \mu(\mathbf{z}) \sum_{i,k,l} S_{j,k}^{-1}(\mathbf{R}') \partial_i S_{k,l}(\mathbf{R}') z_i z_l = T\Delta \sum_{i,k,l} S_{j,k}^{-1}(\mathbf{R}') \partial_i S_{k,l}(\mathbf{R}') S_{i,l}^{-1}(\mathbf{R}')$$

$$= -T\Delta \sum_i \partial_i S_{j,i}^{-1}(\mathbf{R}') = -T\Delta \Gamma_j(\mathbf{R}') \tag{32}$$

where in the last equality, we have used the relation given in Eq. (30). In this way the total term proportional to $\partial_j P_n(\mathbf{R}')$ reads:

$$\sum_j \left\{ -\Delta \left[ S^{-1}(\mathbf{R}') \mathbf{f}_{\mathbf{x}'} \right]_j + T\Delta \Gamma_j(\mathbf{R}') \right\} \partial_j P_n(\mathbf{R}') \tag{33}$$

Finally the term proportional to the second derivative leads to:

$$\Delta T \sum_{i,j} S_{ij}^{-1}(\mathbf{R}') \partial_i \partial_j P_n(\mathbf{R}') = T\Delta \sum_i \partial_i \left[ S_{ij}^{-1}(\mathbf{R}') \partial_j P_n(\mathbf{R}') \right]$$

$$- T \sum_j \left[ \sum_i \partial_i S_{i,j}^{-1}(\mathbf{R}') \right] \partial_j P_n(\mathbf{R}')$$

$$= T\Delta \sum_i \partial_i \left[ S_{ij}^{-1}(\mathbf{R}') \partial_j P_n(\mathbf{R}') \right]$$

$$- T\Delta \sum_j \Gamma_j(\mathbf{R}') \partial_j P_n(\mathbf{R}') \tag{34}$$

By collecting all the terms obtained in Eqs. (24), (33), and (34), all the ones proportional to the "cumbersome gradient" $\Gamma_j(\mathbf{R}')$ cancel out, and, by carrying out the limit $\Delta \to 0$, we obtain the following Fokker-Planck equation:

$$\partial_t P(\mathbf{R}, t) = \sum_j \partial_j \left\{ -\left[S^{-1} f_{\mathbf{R}}\right]_j P(\mathbf{R}, t) + T \sum_i S_{i,j}^{-1} \partial_i P(\mathbf{R}, t) \right\} + \bar{C}(\mathbf{R}) P(\mathbf{R})$$

(35)

where all the terms proportional to $P(\mathbf{R})$ include $C(\mathbf{R})$ and the ones that compensate the expression implied by the total divergence, namely, $\bar{C}(\mathbf{R}) = C(\mathbf{R}) + \sum_j \partial_j \left[S^{-1} f_{\mathbf{R}}\right]_j$. In the above equation, $\bar{C}(\mathbf{R})$ has not been computed explicitly as it should simply vanish because it is determined by the standard property of the Fokker-Planck equation, namely, that the RHS should be a total divergence, so that once integrated overall volume, it guarantees that the normalization of the probability is conserved for any initial probability guess, as a simple consequence that the conditional probability satisfies $\int d[R']^p K(\mathbf{R}'|\mathbf{R}) = 1$ for any $\Delta$ and in particular in the limit $\Delta \to 0$. Therefore we finally obtain the following Fokker-Planck equation with $\bar{C}(\mathbf{R}) = 0$:

$$\partial_t P(\mathbf{R}, t) = \sum_j \partial_j \left\{ -\left[S^{-1}[\mathbf{R}] f_{\mathbf{R}}\right]_j P(\mathbf{R}, t) + T \sum_i \left[S_{j,i}^{-1}[\mathbf{R}] \partial_i P(\mathbf{R}, t)\right] \right\}$$

(36)

that concludes the proof of this section.

## 3.1  A More Efficient Algorithm

The previous Markov chain given in Eq. (17) solves the problem of computing the "cumbersome term" $\Gamma_j(\mathbf{R})$ at the expense of evaluating the matrix $S(\mathbf{R})$ twice for each iteration. In the following we describe another way to obtain the same Fokker-Planck equation, with an iterative scheme requiring only one evaluation of the matrix $S(\mathbf{R})$. This is important in our implementation of the Langevin dynamics, whenever the evaluation of the matrix $S(\mathbf{R})$ requires high computational effort as in computationally demanding applications based on DFT or quantum Monte Carlo. To this purpose the following more convenient iteration scheme defines the new coordinates $\mathbf{R}(t + \Delta)$ not only in terms of $\mathbf{R}(t)$ but also of the previous one $\mathbf{R}(t - \Delta)$. This remains formally a Markov chain in an extended space acting on a $2p-$ dimensional vector $\tilde{\mathbf{R}}_n = [\mathbf{R}_n, \mathbf{R}_{n-1}]$, so that all the results of Markov chains used in the previous section can be applied also in this case. We propose therefore the following iteration scheme:

$$\mathbf{R}(t + \Delta) = \mathbf{R}(t) + \Delta S^{-1}[\mathbf{R}(t)] \mathbf{f}_{\mathbf{R}(t)} + \sqrt{2T\Delta} \mathbf{z}(t)$$

$$-\frac{1}{2} S^{-1}[\mathbf{R}(t)] \{S[\mathbf{R}(t - \Delta)] - S[\mathbf{R}(t)]\} [\mathbf{R}(t - \Delta) - \mathbf{R}(t)]$$

$$\langle z_i(t) z_j(t) \rangle = S_{i,j}^{-1}[\mathbf{R}(t)]$$

(37)

In order to verify that also this Markov chain converges, for $\Delta \to 0$ to the equilibrium distribution given in Eq. (8), it is enough to show that, at the leading order in $\Delta$, the previous iteration scheme is equivalent to the Markov chain in Eq. (17). This is evident by considering that, as in the previous case:

$$(\mathbf{R}(t - \Delta) - \mathbf{R}(t)) = -\sqrt{2T\Delta}\mathbf{z}(t - \Delta) + O(\Delta) \tag{38}$$

Therefore the "cumbersome term" in Eq. (37) comes naturally from simple Taylor expansion:

$$\begin{aligned}
&\{[S(\mathbf{R}(t - \Delta)) - S(\mathbf{R})] (\mathbf{R}(t - \Delta) - \mathbf{R}(t))\}_k \\
&= 2T\Delta \sum_{i,l} \left(\partial_i S_{k,l}(\mathbf{R}(t))\right) z_i(t - \Delta) z_l(t - \Delta) \\
&\simeq 2T\Delta \sum_{i,l} \left(\partial_i S_{k,l}(\mathbf{R})\right) z_i(t) z_l(t)
\end{aligned} \tag{39}$$

Thus the iteration scheme in Eq. (37) is equivalent to the following Markov chain:

$$\begin{aligned}
\mathbf{R}(t + \Delta)_j &= \mathbf{R}(t)_j + \Delta \left(S^{-1}[\mathbf{R}]\mathbf{f_R}\right)_j \\
&\quad -T\Delta \sum_{i,k,l} S_{j,k}^{-1}[\mathbf{R}] \left(\partial_i S_{kl}[\mathbf{R}]\right) z_i(t) z_l(t) + \sqrt{2T\Delta} z_j(t) \\
&\langle z_i(t) z_j(t)\rangle = S_{i,j}^{-1}[\mathbf{R}(t)],
\end{aligned} \tag{40}$$

namely, it coincides with the Markov chain we have considered in the previous section up to order $O(\Delta)$ because one can substitute in Eq. (17) $[S(\mathbf{y}) - S(\mathbf{R})]_{kl}$ with its expansion in $\Delta$: $\sqrt{2T\Delta} \sum_i \partial_i S_{kl} z_i + O(\Delta)$.

Strictly speaking the rightmost equality in Eq. (39) is valid for the associated Fokker-Planck equation where one can substitute the correlator $\langle z_i(t - \Delta) z_l(t - \Delta)\rangle$ with the one $\langle z_i(t) z_l(t)\rangle$, clearly allowed because in this way only a negligible error $O(\Delta)$ is introduced.

We have therefore verified that the Markov chain in Eq. (17) is equivalent to the proposed one in Eq. (40) that can be more efficiently implemented while leading to the same Fokker-Planck equation for $\Delta \to 0$ and therefore to the same equilibrium distribution in Eq. (8).

## 3.2    Renormalization of the Target Temperature

In the case of an harmonic potential $V_H$ with equilibrium position $\mathbf{R}^{eq}$, as defined in Eq. (10), it is readily seen that for $\Delta = 1$, with a matrix $S = H$ independent of $\mathbf{R}$, the Markov chain of Eq. (37) simplifies to:

$$\mathbf{R}(t + \Delta) = \mathbf{R}^{eq} + \sqrt{2T}\mathbf{z}$$

$$\langle z_i z_j \rangle = H_{i,j}^{-1} \tag{41}$$

which corresponds to the direct sampling of the canonical distribution $\exp(-V_H/2T)$, namely, correlation time equal to one and temperature renormalized by a factor two. Thus, in the harmonic case, the present scheme provides a simple scaling of the simulated target temperature $T_{target}$, which turns out to be two times the nominal one $T$ for large time steps $\Delta = 1$ and $S = H$. This consideration suggests that it is possible to apply Eq. (37) and determine the target temperature of the simulation a posteriori as it is often done in the case of second-order Langevin dynamics, by measuring the average squared velocities of the particles. Unfortunately a direct estimation of the simulated temperature is rather difficult in the first-order Langevin dynamics, as the virial theorem usually provides a quite noisy estimate of this quantity and it is also difficult to apply in periodic systems.

## 3.3   Efficiency Gain in Controllable Models

In this section we break our discussion to apply and numerically demonstrate the efficiency of the method. We use different definitions for the matrix $S$ to prove the generality of our framework. In the first, low-dimensional example, the matrix $S$ is derived from geometrical considerations, while in the second case, a dense liquid of dimers interacting with a Lennard-Jones potential, we use a regularized version of the Hessian matrix. A third possibility, which can only be implemented in combination with a quantum Monte Carlo calculation of the forces, will be discussed in Sect. 3.5.

### 3.3.1   Rotating Spring on a Plane

We first consider a simple toy model, a rotating spring lying on a plane. The endpoint $(x, y)$ is subject to a radial harmonic potential of the form

$$U(x, y) = \frac{1}{2}k \left( \sqrt{x^2 + y^2} - a \right)^2 , \tag{42}$$

while being free to rotate around the origin. The configuration's space visited during the dynamics is a circular ring, whose radius is $a$ and width given by the (radial) thermal fluctuation. This is perhaps the simplest model where a strong decoupling of time scale, in this case the vibrational and rotational ones, is present.

Let us define the following matrix $G_\lambda$:

$$G_\lambda = \frac{1}{\lambda} \begin{pmatrix} \frac{x^2 + y^2\lambda}{x^2+y^2} & \frac{xy(1-\lambda)}{x^2+y^2} \\ \frac{xy(1-\lambda)}{x^2+y^2} & \frac{x^2\lambda+y^2}{x^2+y^2} \end{pmatrix} \tag{43}$$
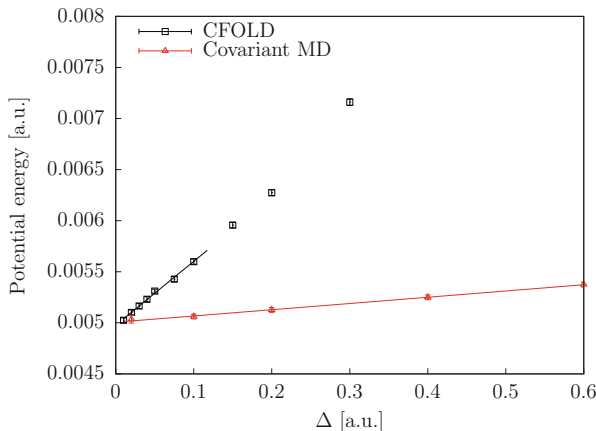
**Fig. 1** Average potential energy as a function of the integration time step $\Delta$. Black points refer to the CFOLD (Eq. (1)), while red points to the improved dynamics (Eq. (37)) with nontrivial metric tensor $S = G_{\lambda=0.1}$, from Eq. (43). The red data series is obtained with the integration scheme of Eq. (37). We use the following parameters, $T = 0.01$, $k = 2$, $a = 1.4$. We see that the new approach greatly alleviates the time step error compared to the standard dynamics while maintaining a very similar rotational diffusion coefficient at fixed $\Delta$ (not shown). Solid lines represent linear fit of the respective data series

where $\lambda$ is a control parameter and $G$ reduces to the identity when $\lambda = 1$. It can be shown, following geometrical considerations, that, if $G_{\lambda}^{-1}$ multiplies the forces, it effectively reduces by a factor $\lambda$ the radial component of the associated displacement. Since this matrix is always positive definite, it represents a good candidate for the dynamics preconditioner. We therefore use $S = G_{\lambda}$ in Eq. (12), with $\lambda < 1$.

From Fig. 1 we see that the preconditioned Langevin dynamics, with nontrivial $S$, results in a better time step error, compared to the standard Langevin dynamics. This demonstrates also in a simple toy model that a large computational gain can be achieved by this framework. In particular the iterative scheme of Eq. (37) removes the time step bias almost completely.

### 3.3.2   A Lennard-Jones Molecular Liquid

The second model is much more realistic, where the force field arises from the following potential:

$$
V = V_{\text{mol}} + V_{LJ} = \sum_{[i,j]_{\text{mol}}}^{N} \frac{1}{2}k(r_{ij} - a_0)^2 + \sum_{[i \neq j]_{\text{mol}}}^{N} 4\epsilon_0 \left[ \left( \frac{\sigma_0}{r_{ij}} \right)^{12} - \left( \frac{\sigma_0}{r_{ij}} \right)^6 \right]
$$

(44)

where $N$ is the total number of atoms, $r_{ij}$ is the distance between atoms $i$ and $j$, and the symbols $[i, j]_{\text{mol}}([i \neq j]_{\text{mol}})$ indicate that the sum includes (excludes) atoms which belong to the same dimer. The force field parameters are chosen to mimic

the radial pair distribution of a dense molecular liquid. In the simulations we fixed $a_0 = 1.4$ a.u., to mimic the equilibrium bond length of the hydrogen molecule (which is nevertheless reduced to $\approx 1.38$ due to the intermolecular interactions), $\sigma_0 = 2$ and $k = 10$, $\epsilon_0 = 0.03$. We set the temperature $T = 0.1$ (with $k_B = 1$) and a particle density given by 0.07 a.u.$^{-3}$. We use periodic boundary conditions.

This dense liquid of dimers is strongly nonharmonic; therefore, its Hessian is not positive definite for any particle configuration. We regularize the Hessian in the following way: we first decompose the $H$ matrix into $H = Q \Lambda Q^T$, where $Q$ contains the eigenvectors (in columns) of $H$ and $\Lambda$ is the diagonal matrix containing the eigenvalues $\lambda_i$, with $i = 1, \cdots, 3N$. Then we regularize the eigenvalues $\lambda_i$ of the Hessian using a smooth function, such that $\lambda_{reg} = u(\lambda)$, with

$$u(\lambda) = \frac{1}{\delta} n((\lambda - \varepsilon)/\tau) + \lambda(1 - n((\lambda - \varepsilon)/\tau)), \qquad (45)$$

where $n(x) = 1/(1+e^x)$ is the Fermi function. This works better than simply cutting off the eigenvalues smaller than $\varepsilon$, because in such a case, the finite difference evaluation of the derivative of $S$ in Eq. (37) could be strongly biased. These parameters are system dependent. We choose $\varepsilon = 1$, $\tau = 2/\delta$. After this procedure we can use this matrix in the preconditioned Langevin equation (Eq. (37)). We notice that much better choices may exist for the hessian regularization.

We again test the time step bias of the different dynamics. We plot in Fig. 2 the results as a function of the dimensionless parameter $\Delta/\tau_{corr}$, where $\tau_{corr}$ is the largest correlation time that can be measured for this system. In our case, one
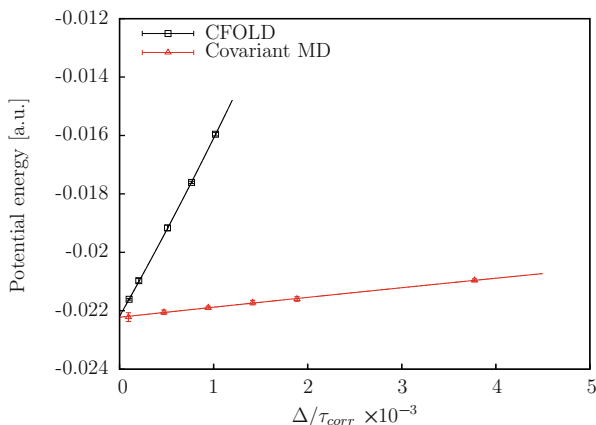


**Fig. 2** Average potential energy of the MD simulations as a function of the integration time step $\Delta$, renormalized with the measured autocorrelation time $\tau_{corr}$. The system consists of 24 particle dimers in a box, interacting via Eq. (44). We compare the CFOLD integrator (black) with the proposed one, by using Eq. (37) (red), using the regularized Hessian matrix. Solid lines represent quadratic fit of the respective data series

of the slowly varying processes is the autocorrelation time of the intermolecular component, $V_{LJ}$, of the total potential. We measure this quantity for the different dynamics, to have an unbiased estimation of the efficiency gain.

## 3.4    Nuclear Quantum Effects

As it is well known, it is possible to simulate nuclear quantum effects at finite temperature by using the classical mapping to a system of $P$ replicas – usually called beads – of the classical system, provided we make two basic assumptions: (i) the temperature is small but not too small that the statistics of nuclei becomes important; (ii) we can use the Born-Oppenheimer approximation in order to decouple the nuclear motion from the electronic one (Ceriotti et al. 2010; Tuckerman et al. 1993). The resulting classical potential, at rescaled temperature $T_P = T \times P$, is defined in terms of the $p = 3NP$ degrees of freedom $\mathbf{R} = \{\mathbf{r}_i^a\}$, for $1 \leq i \leq N$ and $0 \leq a \leq P - 1$, with periodic boundary conditions, i.e., $\mathbf{r}_i^P = \mathbf{r}_i^0$, and can be written as follows:

$$V_P(\mathbf{R}) = \sum_{i=1}^{N} \sum_{a=0}^{P-1} V(\mathbf{r}_i^a) + \frac{m_i T_P^2}{2} \left( \mathbf{r}_i^a - \mathbf{r}_i^{a+1} \right)^2 \tag{46}$$

where $m_i$ is the mass of the $i$th particle in units $k_B = \hbar = 1$. The Hamiltonian (46) describes a classical system of $PN$ particles that are interacting with an Hessian matrix, whose leading large $P$ term do not couple the $3N$ spatial coordinates $\nu, \mu$:

$$K_{a\nu,b\mu} = \delta_{\nu,\mu} \frac{m_i T_P^2}{2} (2\delta_{a,b} - \delta_{a+1,b} - \delta_{b+1,a}) + \frac{1}{2}\delta_{a,b} \left[ \partial_\nu \partial_\mu V(\mathbf{R}^a) \right] \tag{47}$$

Therefore if we neglect the less relevant terms inside the square brackets, coming from the expansion of $V$, this matrix can be readily diagonalized by a Fourier transform, with plane wave eigenvectors $q_{\omega_n}^\nu(a) = C_{\omega_n} \exp(i\omega_n a)$ with $\omega_n = \frac{2\pi}{P}n$, $n = 0, 1, \cdots P - 1$, $C_{\omega_n}$ appropriate constants and eigenvalues $m_i T_P^2(1 - \cos \omega_n)$. The condition number of the above matrix can be readily evaluated and is given by $K_{\text{cond}} = P^2/\pi$ because the maximum eigenvalue is obtained for $\omega_n = \pi$ and the minimum non-zero one for $\omega_n = \frac{2\pi}{P}$. Thus it is clear that it is possible to accelerate the dynamics by a factor $\simeq P^2$ using for $S$ the harmonic matrix given in Eq. (47), with a reasonable approximation for the term depending on $V$ because this term is almost negligible, i.e., finite, for $P \rightarrow \infty$. We remark therefore that, also for the simulation of nuclear quantum effects, it is extremely important to use an appropriate acceleration matrix in order to avoid extremely large correlation times.

## 3.5     Accelerated Molecular Dynamics Within QMC Applications

This section concerns the specific application of this LDE in QMC. While Eq. (37) can be applied already as it is when using exactly computed ionic forces $\mathbf{f}$ (from either force-fields or DFT based calculations), one additional step is required in the case of QMC, where forces are given with a certain statistical error.

In the simplest variational quantum Monte Carlo approach (Attaccalite and Sorella 2008; Mazzola et al. 2014; Zen et al. 2015) (VMC), it is possible to perform finite temperature ab initio simulation of classical particles interacting via the Born-Oppenheimer energy surface, through a quantum mechanical variational optimization of a correlated electronic wave function $\Psi_{\alpha}$, which contains several parameters $\{\alpha\}$:

$$V(\mathbf{R}) = \text{Min}_{\alpha} \frac{\langle \Psi_{\alpha} | H_{\mathbf{R}} | \Psi_{\alpha} \rangle}{\langle \Psi_{\alpha} | \Psi_{\alpha} \rangle} \tag{48}$$

where $H_{\mathbf{R}}$ is the full-many body electronic Hamiltonian with Coulomb interaction, at fixed atomic positions. Being a statistical method, the forces are affected by a finite statistical error in QMC (Attaccalite and Sorella 2008) and a very useful matrix naturally emerges – the covariance matrix – describing the correlations between the force noise components:

$$Cov(\mathbf{f}) = \langle\langle f_i(\mathbf{R}) f_j(\mathbf{R}) \rangle\rangle - \langle\langle f_i(\mathbf{R}) \rangle\rangle\langle\langle f_j(\mathbf{R}) \rangle\rangle \tag{49}$$

where $\langle\langle \ \rangle\rangle$ indicates a statistical average over a given number of samples at fixed atomic positions.

In this QMC approach, we use $S = Cov(f)$, where $Cov(f)$ is the correlation matrix corresponding to the statistical fluctuations – i.e., the error bars – of the nuclear forces. In this case it is also very simple to correct for the extra noise given by the QMC forces (Luo et al. 2014). This is achieved in a very simple way, just by changing the temperature $T$ used in the dynamics for the correct simulation at a given target temperature $T_{\text{target}}$, simply as follows:

$$2T\Delta = 2T_{\text{target}}\Delta - \Delta^2. \tag{50}$$

that is possible for $\Delta < 2T_{\text{target}}$ (notice that for this particular choice of $S$ the time of the dynamics has the unusual dimension of an energy).

In order to show that this technique is extremely convenient, even within a computationally demanding method such as QMC, we present in Fig. 3 a molecular dynamics simulation of a mixture containing 118 hydrogens and 10 helium, at Jupiter interior conditions, with an accurate basis set. It is remarkable that, within the same simulation, different phases can be reached in a short simulation time, clearly showing the efficiency of the method and its capability to identify the metal-insulator transition/crossover which is of paramount importance for the understanding of the structure of giant planets, such as Jupiter and Saturn.
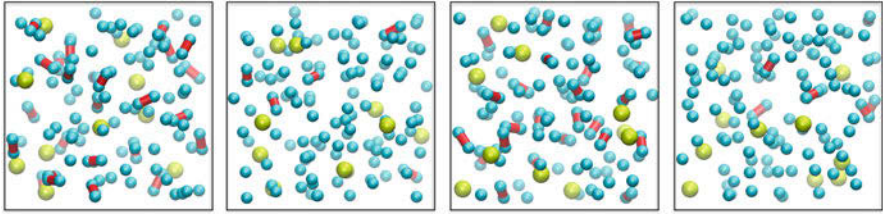
**Fig. 3** Snapshot of an H-He mixture simulation, a NVE molecular dynamics driven by true QMC nuclear forces (Mazzola et al. 2018). The cyan, yellow, and red colors represent the 108 H, 10 He, and H-molecule bonds (H2), respectively. The temperature is 6000 K and the density 0.53 g/cm$^3$ (P = 54 Gpa), i.e., near to the liquid-liquid metal-insulator transition. The occurrence of these fluctuations between qualitatively different phases indicates that a metal-insulator insulator transition or at least a sharp crossover could occur within this pressure range. This has a clear impact for the equation of state of large planets, such as Jupiter, and is important in view of the recent data reported from the NASA JUNO mission

## 4    Conclusions

Within the framework of first-order Langevin dynamics, it is possible to define a method that, for classical dynamics, solves the problem of working with very different time scales, by using an acceleration matrix S, suitably chosen, that generates an efficient dynamics where all the time scales become equal. Roughly speaking in the period when a fast molecular oscillation occurs also a long distance rearrangement of the electronic structure takes place (e.g., weak molecular binding). From the theoretical side, it is shown that, with the present method, if compared with the standard Langevin dynamics without S, a gain in efficiency of the order of the condition number K of the Hessian matrix (or the square of the condition number of the dynamical matrix) is achieved at least at low temperatures where the harmonic approximation holds. Indeed, provided S is chosen to be equal to the Hessian matrix, the proposed algorithm generates essentially an independent configuration for each step of molecular dynamics, as the time step $\Delta$ can be as large as the correlation time, short and common for all energy scales.

This gain in efficiency is remarkable even if compared with state-of-the-art second-order Newtonian dynamics (Ceriotti et al. 2009; Mouhat et al. 2017) because, in the latter cases, only a gain of order K is possible with present algorithms. Considering that $K$ can be as large as 100 in the water dimer and is typically growing with the system size, the potential of this method is easily understood. For a non-molecular solid, the matrix S can be chosen independent from the atomic positions, and in this case the method coincides with the standard method of acceleration (Parisi 1984). The important result of this work is that the dependence of the matrix S from the atomic positions can be taken formally into account at no cost. This is extremely important because, even at very low temperatures, there exist instances of solids where the molecules do not freeze (therefore the elastic constant matrix K depends explicitly on the instantaneous

orientation of the molecules). A simple example is the low-pressure phase I of hydrogen, which consists of freely rotating H2 molecules centered on a hexagonal close-packed lattice (Loubeyre et al. 1996). This is indeed the generic case at finite temperature, as, for instance, in liquid water, it is impossible to speed up the molecular dynamics without considering, in the acceleration matrix S, the orientation of all the water molecules, because their mutual interaction strongly depends on the hydrogen bonds network. It would be extremely interesting to apply our technique within DFT because it is well known that extremely long simulations are necessary to obtain reliable equilibrium properties in water, the autocorrelation time for Newtonian MD with stochastic velocity rescaling thermostat being around 1 ps (Grossman et al. 2004; Dawson and Gygi 2018). In this case the best choice for S has not been determined yet, but it is likely that, by taking for S a simple Hessian corresponding to independent water molecules, a good speed up should be achieved. So far this technique has been implemented within the ab initio molecular dynamics tool provided in the TurboRVB QMC package and significative performances have been achieved in several systems, from nanotubes (Varsano et al. 2017) to hydrogen-helium mixtures (Mazzola et al. 2018) and structural optimization of graphene (Sorella et al. 2018), but several other applications are possible, especially considering that this useful and simple scheme can be applied even within more conventional ab initio simulation methods based on DFT. It would be also extremely important to generalize this technique within the second-order Newton dynamics, because the time has clear physical meaning in this case. Indeed, the generalization of the acceleration within the Newtonian dynamics with friction has not been carried out so far, and this should be worthwhile for the future development of the method. In principle, with this technique, one could carry out ab initio molecular dynamics, where only the longest, more interesting, time scales remain meaningful, all the other being rescaled.

# References

Allen M, Tildesley D (1987) Computer simulation of liquids. Oxford University Press, Oxford

Attaccalite C, Sorella S (2008) Stable liquid hydrogen at high pressure by a novel ab initio molecular-dynamics calculation. Phys Rev Lett 100:114501. https://doi.org/10.1103/PhysRevLett.100.114501

Bennett CH (1975) Mass tensor molecular dynamics. J Comput Phys 19:267

Ceriotti M, Bussi G, Parrinello M (2009) Langevin equation with colored noise for constant-temperature molecular dynamics simulations. Phys Rev Lett 102(2):20601

Ceriotti M, Parrinello M, Markland TE, Manolopoulos DE (2010) Efficient stochastic thermostatting of path integral molecular dynamics. J Chem Phys 133:124104

Dawson W, Gygi F (2018) Equilibration and analysis of first-principles molecular dynamics simulations of water. J Chem Phys 148(12):124501. https://doi.org/10.1063/1.5018116

Dykstra CE, Frenking G, Kim KS, Scuseria G (2005) Theory and applications of computational chemistry: the first forty years. Elsevier, Amsterdam

Foulkes WMC, Mitas L, Needs RJ, Rajagopal G (2001) Quantum monte carlo simulations of solids. Rev Mod Phys 73(1):33–83. https://doi.org/10.1103/RevModPhys.73.33

Grossman JC, Schwegler E, Draeger EW, Gygi F, Galli G (2004) Towards an assessment of the accuracy of density functional theory for first principles simulations of water. J Chem Phys 120:300

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140(4A):A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Kühne TD, Krack M, Parrinello M (2009) Static and dynamical properties of liquid water from first principles by a novel car-parrinello-like approach. J Chem Theory Comput 5(2):235–241. https://doi.org/10.1021/ct800417q

Loubeyre P, LeToullec R, Hausermann D, Hanfland M, Hemley RJ, Mao HK, Finger LW (1996) X-ray diffraction and equation of state of hydrogen at megabar pressures. Nature 383:702 EP –. https://doi.org/10.1038/383702a0

Luo Y, Zen A, Sorella S (2014) Abinitio molecular dynamics with noisy forces: validating the quantum monte carlo approach with benchmark calculations of molecular vibrational properties. J Chem Phys 141(19):194112. https://doi.org/10.1063/1.4901430

Mazzola G, Sorella S (2017) Accelerating ab initio molecular dynamics and probing the weak dispersive forces in dense liquid hydrogen. Phys Rev Lett 118:015703. https://doi.org/10.1103/PhysRevLett.118.015703

Mazzola G, Zen A, Sorella S (2012) Finite-temperature electronic simulations without the born-oppenheimer constraint. J Chem Phys 137(13):134112. https://doi.org/10.1063/1.4755992

Mazzola G, Yunoki S, Sorella S (2014) Unexpectedly high pressure for molecular dissociation in liquid hydrogen by electronic simulation. Nat Commun 5:3487. https://doi.org/10.1038/ncomms4487

Mazzola G, Helled R, Sorella S (2018) Phase diagram of hydrogen and a hydrogen-helium mixture at planetary conditions by quantum monte carlo simulations. Phys Rev Lett 120:025701. https://doi.org/10.1103/PhysRevLett.120.025701

Mouhat F, Sorella S, Vuilleumier R, Saitta AM, Casula M (2017) Fully quantum description of the zundel ion: combining variational quantum monte carlo with path integral langevin dynamics. J Chem Theory Comput 13(6):2400–2417. pMID: 28441484. https://doi.org/10.1021/acs.jctc.7b00017

Parisi G (1984) Progress in Gauge field theory. NATO ASI series. Springer, Boston

Risken H (1996) The Fokker-Planck equation: methods of solution and applications. Springer, Berlin

Scheraga HA, Khalili M, Liwo A (2007) Protein-folding dynamics: overview of molecular simulation techniques. Annu Rev Phys Chem 58:57–83

Sorella S, Seki K, Brovko OO, Shirakawa T, Miyakoshi S, Yunoki S, Tosatti E (2018) Structural dimerization, electron correlations, and topological gap opening in isotropically strained graphene. Phys. Rev. Lett. 121:066402

Tassone F, Mauri F, Car R (1994) Acceleration schemes for ab initio molecular-dynamics simulations and electronic structure calculations. Phys Rev B 50:10561

Tsuchida E (2015) Ab initio mass tensor molecular dynamics. J Chem Phys 134:044112

Tuckerman M (2010) Statistical mechanics: theory and molecular simulation. Oxford University Press, Oxford

Tuckerman ME, Berne BJ, Martyna GJ, Klein ML (1993) Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals. J Chem Phys 99(4):2796–2808

Varsano D, Sorella S, Sangalli D, Barborini M, Corni S, Molinari E, Rontani M (2017) Carbon nanotubes as excitonic insulators. Nat Commun 8:1461

Zen A, Luo Y, Mazzola G, Guidoni L, Sorella S (2015) Ab initio molecular dynamics simulation of liquid water by quantum monte carlo. J Chem Phys 142(14):144111

# Metadynamics: A Unified Framework for Accelerating Rare Events and Sampling Thermodynamics and Kinetics

**27**

Giovanni Bussi, Alessandro Laio, and Pratyush Tiwary

## Contents

---

G. Bussi
SISSA, Trieste, Italy
e-mail: bussi@sissa.it

A. Laio (✉)
SISSA, Trieste, Italy
International Centre for Theoretical Physics (ICTP), Trieste, Italy
e-mail: laio@sissa.it

P. Tiwary
Department of Chemistry and Biochemistry and Institute for Physical Science and Technology,
University of Maryland, College Park, MD, USA
e-mail: ptiwary@umd.edu; pratyush.tiwary@gmail.com

**Abstract**

Metadynamics is an enhanced sampling algorithm in which the normal evolution of the system is biased by a history-dependent potential constructed as a sum of Gaussians centered along the trajectory followed by a suitably chosen set of collective variables. The sum of Gaussians forces the system to escape from local free energy minima and is used to iteratively build an estimator of the free energy. This original idea has been developed and improved over the years in several variants, which nowadays allow addressing in a unified framework some of the most important tasks of molecular simulations: computing the free energy as a function of the collective variables, accelerating rare events, and estimating unbiased kinetic rate constants. This chapter provides a survey of the many formulations of metadynamics with an emphasis on the underlying theoretical concepts and some hints on the appropriate manner of using this approach for solving complicated real-world problems.

## 1    Introduction

The metadynamics method was introduced in 2002 (Laio and Parrinello 2002), and it has rapidly become popular for the study of thermodynamics and kinetics in molecular systems that are hard to sample even with the best available computing resources. This algorithm, like umbrella sampling, requires the preliminary identification of a set of collective variables (CVs) that are assumed to describe the process of interest. A bias depending on these CVs is then constructed iteratively and dynamically as a sum of Gaussians centered along the trajectory followed by the CVs. The longer the dynamics is continued, the more Gaussians are added. Therefore, the bias is time dependent, and it never stops changing. These Gaussians discourage the system from revisiting the same spot in CV space and encourage an efficient exploration of the free energy surface. Indeed, if the dynamics is started in a free energy minimum, the Gaussian potentials accumulate and fill this minimum, until the system moves to a close-by free energy minimum. This allows the system to migrate from well to well. In the original non-tempered form of metadynamics, when all the wells are filled with Gaussians, the dynamics in the CV space becomes diffusive, and the bias potential of metadynamics resembles the ideal bias of umbrella sampling, namely, the negative of the free energy.

Biasing the dynamics with Gaussians is an idea that was introduced well before the 2002 paper. The taboo search method (Cvijovic and Klinowski 1995) and, in the context of molecular dynamics, the local elevation method (Huber et al. 1994) are based on the same idea. Also the idea of using a history-dependent bias to estimate thermodynamic quantities, in particular the density of states, was introduced before, in the Wang and Landau algorithm (Wang and Landau 2001). The main novelty introduced by metadynamics is exploiting these ideas together with dimensionality reduction, in which the dynamics is considered important and biased only in the CV space, in the spirit of the work by Kevrekidis (Theodoropoulos et al. 2000;

Kevrekidis et al. 2004). In practical terms, this allows using the framework of taboo search for biasing efficiently more than one CV at the same time, still obtaining a meaningful free energy estimate. In 2002, the most popular approach for computing the free energy was thermodynamic integration (Carter et al. 1989), which allows using only one CV at a time. The possibility offered by metadynamics of obtaining "on-the-fly" free energy landscapes in two or three dimensions was extremely important in the field and is possibly one of the main reasons of the success of the algorithm.

A second important advantage of metadynamics is that the method can be proficiently used also for accelerating rare events and studying the reaction mechanism of complex processes. Since the history-dependent potential iteratively compensates the underlying free energy, a system evolved with metadynamics tends to escape from any free energy minimum via the lowest free energy saddle point. This makes metadynamics a rather flexible tool that can be used not only to compute efficiently the free energy but also to explore new reaction pathways, accelerate the observation of rare events, and even compute kinetic rate constants in a relatively straightforward manner. If the CVs are chosen sensibly, the system will quickly find its way over the lowest free energy saddle point and evolve over the next minimum as it would eventually do in a very long molecular dynamics simulation. This flexibility is reflected in the numerous contexts in which this method has been applied so far, including solid state and material science, crystal structure prediction, biophysics and chemistry. A final advantage of metadynamics is its inherent self-diagnostic capability. Indeed, the dynamics of the CVs under the action of the bias can provide useful signs to alert user whether the reconstructed free energy/kinetics can be trusted or not.

The key assumption of the method is that the time-dependent potential defined by the sum of Gaussians deposited up to time $t$ provides an unbiased estimate of the free energy in the region explored during the dynamics. This property, which does not follow from any ordinary thermodynamic identity, such as umbrella sampling (Patey and Valleau 1975), was postulated on a heuristic basis in Laio and Parrinello (2002) and afterward verified empirically in several systems of increasing complexity. Successively (Bussi et al. 2006b), it was shown that this property derives from rather general principles and can be demonstrated rigorously for systems in which the dynamics of the CVs is adiabatically separated from the other degrees of freedom. More recently, a modified version of the algorithm was developed in which convergence can be proved explicitly and rigorously for a generic dynamics (Barducci et al. 2008). In this approach, called well-tempered metadynamics, the rate at which the history-dependent potential is modified gradually decreases during the simulation according to a schedule that allows fluctuations in the potential to be damped out. This is done by introducing an extra parameter that controls this damping and that can be used to interpolate between unbiased molecular dynamics and non-tempered metadynamics. Notably, the modified algorithm only samples a low free energy portion of the CV space and in the long limit corresponds to altering the temperature of the biased CVs alone. A rigorous proof of convergence of the algorithm in this formulation is presented in Dama et al. (2014).

## 2    The Basic Algorithm

Consider a system described by a set of coordinates $x$ and a potential $V(x)$ evolving under the action of a dynamics, which could be, for instance, Langevin, Newtonian (under the action of a thermostat), or Monte Carlo, whose equilibrium distribution is canonical at a temperature $T$. The set of coordinates $x$ may include ordinary atomic positions, but also electronic coordinates, as in Car-Parrinello molecular dynamics (Car and Parrinello 1985), or any other auxiliary variables. We aim at exploring the properties of the system as a function of a finite number of CVs $S_\alpha(x)$, $\alpha = 1 \ldots d$ where $d$ is a small number. The CVs can be any explicit function of $x$ such as, for example, a distance, a coordination number, or the gyration radius of a subset of atoms. The free energy $F(s)$ is given by

$$F(s) = -\frac{1}{\beta} \log \left( \int dx \exp\left(-\beta V(x)\right) \delta\left(s - S(x)\right) \right). \tag{1}$$

where $\beta = \frac{1}{k_B T}$ is the inverse temperature. In Eq. 1 (and in the following), capital $S$ is used for denoting the function of the coordinates $S(x)$, while lower case $s$ is used for denoting the value of the CVs.

In the simplest molecular dynamics implementation of metadynamics, the external potential acting on the system at time $t$ is given by

$$V_G(S(x), t) = w \sum_{\substack{t' = \tau_G, 2\tau_G, \cdots \\ t' < t}} \exp\left( -\frac{\left(S(x) - s\left(t'\right)\right)^2}{2\delta s^2} \right) \tag{2}$$

where $s(t) = S(x(t))$ is the value taken by the CV at time $t$. This specific variant of the algorithm is referred to as "direct metadynamics," to distinguish it from previous variants of the algorithm where the CV dynamics was performed separately from that of the rest of the system either using time-dependent restraints (Laio and Parrinello 2002) or an extended Lagrangian formulation (Iannuzzi et al. 2003). Three parameters enter the definition of the $V_G$:

1. The Gaussian height $w$.
2. The Gaussian width $\delta s$.
3. The frequency $\tau_G$ at which the Gaussians are added.

The first two parameters define the shape of the extra bias added at time intervals of $\tau_G$. As it will be discussed in detail in the following, if the Gaussians are large, the free energy surface will be explored at a fast pace, but the reconstructed profile will be characterized by large fluctuations.

As an example, consider the system in Fig. 1. Metadynamics is performed on the one-dimensional potential with three minima represented in the lower panel.

The system evolves through an overdamped Langevin equation (Risken 1989) with time step 1, diffusion coefficient D = 0.005, and at a temperature of 1. The dynamics is started from the central minimum. Without any bias, the system would escape from this minimum with low probability, since the barriers separating it from the other minima are large with respect to the thermal energy. The upper panel shows the trajectory followed by the system under the action of a metadynamics bias. A Gaussian of width $\delta s = 0.4$ and height $w = 0.3$ is added every 300 steps. After ~20 Gaussians, the central minimum is filled, and the system escapes from the well through the lowest saddle point (blue lines in Fig. 1). The second well is filled after ~70 Gaussians (red lines). The second highest saddle point is reached after ~100 Gaussians, and the full free energy surface is filled with a total of ~180 Gaussians (orange lines). After that time, the motion of the system becomes diffusive and unbound in the region of CV space between ~−5 and 5. The metadynamics potential $V_G$ (Eq. 2) is represented at different times in Fig. 1, middle panel. The lower panel of Fig. 1 reports the sum of $V_G$ and of the external potential (thick black line). Clearly, as the simulation proceeds, $V_G$ iteratively compensates the underlying potential.

This provides a hint of the two different manners in which metadynamics can be used:

- It can be used to escape free energy minima (Laio and Parrinello 2002), namely, to find the lowest free energy saddle point out of a free energy minimum. In this case, metadynamics could be stopped as soon as the walker exits from the minimum and starts exploring a new free energy minimum. In Fig. 1, this happens after ~20 Gaussians are placed.
- It can be used to explore a region in the CV space including two or more free energy minima and reconstruct the free energy surface. In this case, the simulation could be stopped when the motion of the walker becomes diffusive in this region. In Fig. 1, this happens after ~180 Gaussians are placed.

The basic assumption of metadynamics is that $V_G(s, t)$ defined in Eq. 2 after a sufficiently long time provides an estimate of the underlying free energy:

$$V_G(s, t) \sim -F(s) + C(t). \tag{3}$$

where $C(t)$ depends on time but not on the collective variables $s$. This equation states that an equilibrium quantity, namely, the free energy, can be estimated by a non-equilibrium dynamics in which the bias potential is changed continuously, every time a new Gaussian is added. Equation 3 was postulated heuristically in Laio and Parrinello (2002), based on the behavior observed in model systems. For instance, in the example of Fig. 1, it is clear that the sum of $F$ and $V_G$ after ~180 Gaussians is approximately a constant. For an atomistic system in which the potential depends on several coordinates, the free energy is the result of a complex dimensional reduction; Eq. 3 can be qualitatively understood in the limit of very small $w$. In this limit, $V_G(s, t)$ varies slowly, and the probability to
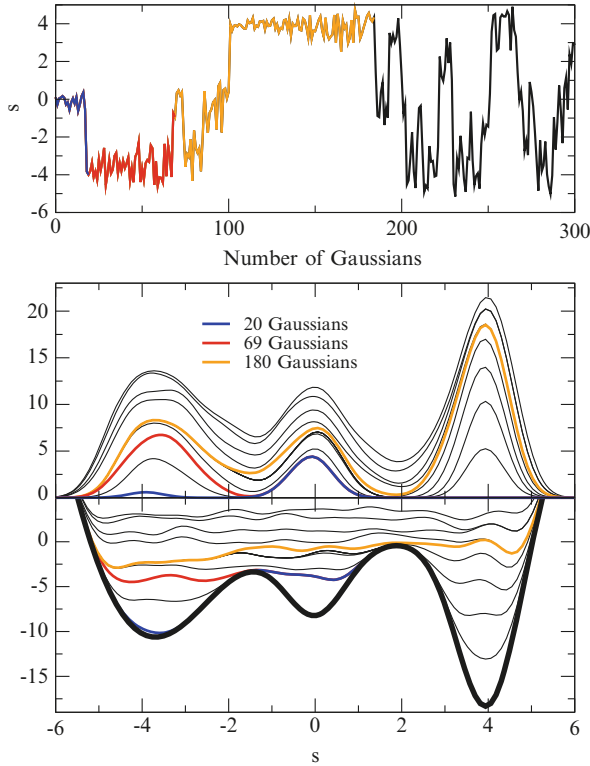
**Fig. 1** Upper panel: trajectory of a one-dimensional system evolved by a Langevin equation on the three-minima potential represented in the lower panel. The dynamics is biased with a metadynamics potential $V_G$ as defined by Eq. 2. The parameters are $\delta s = 0.4$, $w = 0.3$, and $\tau_G = 300$. Middle panel: time evolution of the metadynamics bias potential $V_G$. Blue line, $V_G$ as when the first minimum is filled and the system escapes to the second minimum; red line, $V_G$ as when also the second minimum is filled; orange line, $V_G$ when the entire profile is filled and the dynamics becomes diffusive. Lower panel: time evolution of the sum of the metadynamics potential $V_G$ and of the external potential, represented as a thick black line. (After Laio and Gervasio 2008)

observe $s$ is always approximately proportional to $\exp\left[-\frac{1}{T}\left(F(s) + V_G(s,t)\right)\right]$. If the function $F(s) + V_G(s,t)$ has some local minimum, $s$ will preferentially be localized in this minimum. It is therefore likely that some Gaussians will be added there until the minimum is filled. Let us consider instead the case in which $F(s) + C(t) \sim -V_G(s,t)$ in a region $\Omega(s)$. The probability will be approximately constant in this region, and the new Gaussians will be placed in a random location. Hence, if $w \to 0$, the only corrugations in the free energy that are not flattened by the dynamics will be of the order of the size of the newly added Gaussians. If the dynamics is continued keeping the value of $w$ fixed, more and more Gaussians will be added, and corrugations of a size of the order of $w$ will appear in different locations.

From this qualitative discussion, we can draw a few important observations, which, as we will show, are at the basis of all the approaches which allow obtaining from metadynamics an estimate of the free energy.

- The bias potential defined in Eq. 2 at finite $w$ provides only an approximate estimate of the free energy, since every new Gaussian adds a corrugation to the profile.
- When the dynamics is diffusive in a region of CV space, the corrugations appear randomly in different positions (see Fig. 1). A more accurate estimate of the free energy can be obtained by taking the time average of the bias potential. This manner of estimating the free energy will be discussed in Sect. 3.
- An alternative manner of obtaining an accurate estimate of the free energy is reducing iteratively the value of $w$, in such a way that the size of the corrugations becomes smaller and smaller. In Sect. 4, we will show that this simple idea can be embedded in a powerful conceptual framework, called well-tempered metadynamics, which allows estimating the free energy *exactly*.

## 3    The Bias Potential as a Dynamic Variable

We first describe a procedure for estimating the free energy based on taking the time average of the bias potential defined in Eq. 2. In order to understand in more detail why this can be done, let us consider the case in which the system evolves following an overdamped Langevin dynamics:

$$dx = -\beta D \nabla V dt + \sqrt{2D} dW(t) \tag{4}$$

where $dW(t)$ is a (possibly multivariate) Wiener process and $D$ is the diffusion coefficient. The evolution of this system under the action of metadynamics is modeled adding a history-dependent term:

$$dx = -\beta D \nabla \left( V(x) + \int_0^t dt' g\left( S(x), s(t') \right) \right) dt + \sqrt{2D} dW(t) \tag{5}$$

where $g(s, s')$ is a kernel that specifies how fast the metadynamics potential changes. In the normal implementation, $g$ is a Gaussian of width $\delta s$ and height $w/\tau_G$ (see Eq. 2):

$$g(s, s') = \frac{w}{\tau_G} \exp\left( -\frac{(s - s')^2}{2\delta s^2} \right)$$

Equation 5 describes a non-Markovian process in coordinate space. In fact, the forces acting on the CVs depend explicitly on their history. Due to this non-Markovian nature, it is not clear if, and in which sense, the system can reach a

stationary state. In Bussi et al. (2006b) a formalism was introduced which allows to map this history-dependent evolution into a Markovian process in which the history-dependent potential itself is considered a dynamic variable. Indeed equation 5 can be written as

$$dV_G\,(s,t) = g\,(s - s\,(t))\,dt \tag{6}$$

$$dx = -\beta D\nabla\left(V(x) + \int ds\delta(s - S(x))V_G(s,t)\right)dt + \sqrt{2D}dW(t) \tag{7}$$

These equations are fully Markovian, i.e., the state of the system at time $t + dt$ depends only on the state of the system at time $t$, *if one assumes that the state of the system is not defined by the coordinates x alone but by the coordinates and the bias potential $V_G\,(s)$.* In other words, the bias potential should be considered a *dynamic variable*, whose evolution is coupled to the evolution of the coordinates as implicit in Eq. 7. Under the action of the dynamics, the configuration of the system $(x, V_G)$ follows a trajectory and samples a probability distribution $P\,(x, [V_G])$ where the $P$ is an ordinary function of the coordinates $x$ and a functional of $V_G\,(s)$.

In Bussi et al. (2006b), it was shown that if the dynamics of the CVs is adiabatically decoupled from the dynamics of the other coordinates, in the limit of long time $P$ converges to a Gaussian in functional space, centered on the negative of the free energy $-F\,(s)$. The expected deviation of $V_G\,(s,t)$ from $-F\,(s) + C(t)$, $\varepsilon = \sqrt{\langle(V_G\,(s) + F\,(s))^2\rangle}$, can be estimated explicitly. The exact expression of the error is rather complex and depends on the dimensionality and on the boundary conditions in CV space. Approximately, for a one-dimensional metadynamics in a region of size $S$ (Laio et al. 2005),

$$\varepsilon \sim C_d\sqrt{\frac{Sw\delta s}{\beta D\tau_G}} \tag{8}$$

where $C_d$ is a constant that depends only on the dimensionality. Thus, $V_G$ deviates from $-F\,(s) + C$ by fluctuations that grow with the square root of the Gaussian height $w$, with the Gaussian width $\delta s$, and with the inverse of the diffusion coefficient $D$.

This is illustrated in Fig. 2 in which the results obtained integrating numerically Eq. 5 for four different profiles $F\,(s)$ are shown. The average value of $V_G\,(s,t) + F\,(s)$ is represented as a continuous line in all the four profiles and is constant in all the explored region. The standard deviation of $V_G$, approximately estimated by Eq. 8, is independent on the free energy profile.

These examples illustrate clearly that, if metadynamics is used to compute the free energy, the history-dependent potential at a single time $V_G\,(s,t)$ *should not be used as a free energy estimator*, even if the dynamics of the CVs is adiabatically decoupled. Indeed, $V_G$ is by construction affected by statistical errors. However, as suggested by Eq. 7, $V_G$ should be considered a dynamics variables, and all the
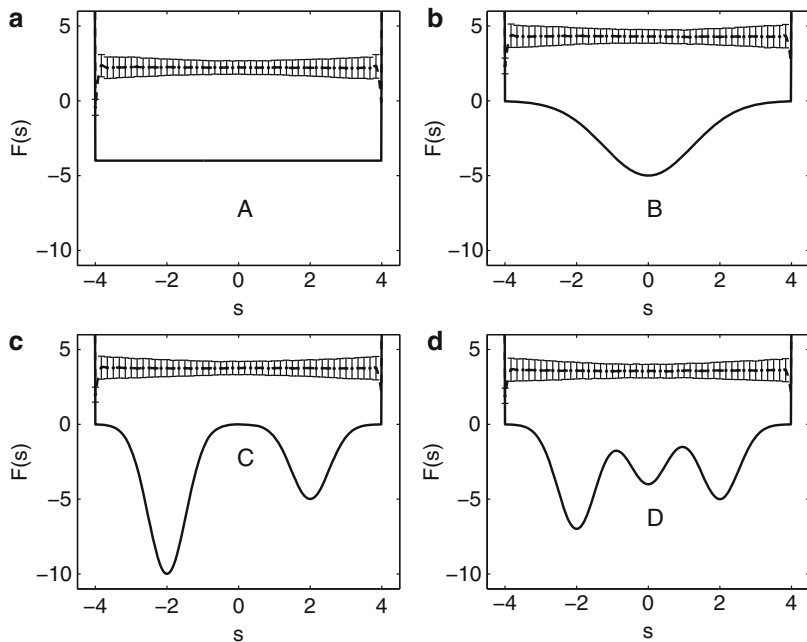
**Fig. 2** Metadynamics results for four different free energy profiles. A, $F(s) = -4$; B, $F(s) = -5\exp\left(-\left(\frac{s}{1.75}\right)^2\right)$; C, $F(s) = -5\exp\left(-\left(\frac{s-2}{0.75}\right)^2\right) - 10\exp\left(-\left(\frac{s+2}{0.75}\right)^2\right)$; D, $F(s) = -5\exp\left(-\left(\frac{s-2}{0.75}\right)^2\right) - 4\exp\left(-\left(\frac{s}{0.75}\right)^2\right) - 7\exp\left(-\left(\frac{s+2}{0.75}\right)^2\right)$. The average $\langle F(s) + V_G(s, t)\rangle$ computed over 1000 independent trajectories is represented as a dashed line, with the error bar estimated as standard deviations. The metadynamics parameters are $\delta s = 0.1$ and $w/\tau_G = 4 \times 10^{-4}$. We also have $D = 0.0005$ and $T = 1$, and the CVs satisfy reflecting boundary conditions in a region of length 8. (After Laio et al. 2005)

profiles $V_G(s, t)$ are equally reliable estimates of the free energy up to a constant. Thus, as first proposed in Micheletti et al. (2004), after the entire free energy landscape has been filled, the best possible estimate of the free energy is the time average of all the profiles. More precisely, if $t_F$ is the time at which the CV starts diffusing in all the relevant region (possibly fixed by the appropriate boundary conditions, see Sect. 3.1) and $t_{\text{tot}}$ is the total simulation time, the best estimate of the free energy is

$$- \bar{V}_G(s) = -\frac{1}{t_{\text{tot}} - t_F} \int_{t_F}^{t_{\text{tot}}} V_G(s, t)\, dt \qquad (9)$$

Similarly to an average of a standard observable in ordinary molecular dynamics or Monte Carlo, the standard deviation of $-\bar{V}_G$ from the free energy decays to zero with the square root of $t_{\text{tot}} - t_F$. The prefactor is determined by the autocorrelation time $\tau_{\text{corr}}$ of the history-dependent potential:

$$\epsilon^2(s) = \left( \bar{V}_G(s) + F(s) \right)^2 \sim \frac{\tau_{\text{corr}}}{t_{\text{tot}} - t_F} \qquad (10)$$

The error on the free energy estimated by Eq. 9 is conveniently evaluated by block analysis. For example, using $N_B$ blocks, one divides the time interval $[t_F, t_{\text{tot}}]$ in $N_B$ nonoverlapping intervals of the same size. One then computes the average of $V_G(s, t)$ restricted to each block. Since the free energy is always defined modulo an irrelevant constant, one chooses this constant in such a way that the average on the domain in which metadynamics is performed of all the block averages is equal to zero. Denoting by $\bar{V}_G^\alpha$ with $\alpha = 1, \dots, N_B$ the averages on the blocks, the error is then estimated as

$$\epsilon^2(s) = \frac{1}{N_B^2} \sum_{\alpha=1}^{N_B} (\bar{V}_G^\alpha(s) - \bar{V}_G(s))^2 \qquad (11)$$

Like in normal block analysis, one should then repeat this calculation with different $N_B$-s and consider the error estimate meaningful if the result is approximately invariant with respect to this choice.

It is important to remark that Eqs. 9 and 10 have been demonstrated rigorously only if the dynamics is adiabatically decoupled in the CVs (Bussi et al. 2006b). However, these equations have been proved to hold numerically in systems in which this condition is violated, in particular a Ising model in which metadynamics is performed using as a CV the magnetization (Crespo et al. 2010) and a peptide in water solution (Marinelli et al. 2009). We are not aware of a molecular system in which a systematic deviation of the estimator in Eq. 9 from the free energy has been reported.

## 3.1    Boundary Conditions

In order to take the average over different profiles in Eq. 9, it is necessary that the dynamics of the CV is bound in a finite region of CV space, in such a way that the history-dependent potential can reach a stationary state. This can be achieved by restraining the dynamics in the region of choice by a suitable external potential. For example, if one wants to estimate the free energy profile for a distance smaller than 5 Å, one can add a harmonic potential acting on the CV when it becomes larger than 5. However, introducing a potential restraining the dynamics in a finite region is not sufficient for reaching a stationary state for the dynamics of $V_G$. Indeed, metadynamics simulations typically take advantage of a finite Gaussian width to fill the free energy surface quickly. Adding more and more Gaussians can induce systematic errors at the boundaries of CV space (whether natural or artificially imposed): these errors are due to the fact that a sum of finite width Gaussians cannot accurately reproduce the free energy profile where the derivative of the free energy is very high (or infinite, in the case of a sharp boundary). At the beginning of the

simulation, these errors are small and are usually overlooked, but at long times they can become important, preventing the dynamics from reaching a stationary condition. These systematic errors close to the CV boundary can be eliminated in at least three different manners.

The first (and simplest) procedure consists in setting the force deriving from the history-dependent potential equal to zero beyond the boundary (Baftizadeh et al. 2012). If, for example, a restraining potential is active for $s > s_w$, one still updates the history-dependent potential according to Eq. 2, but one then sets to zero the forces coming from $V_G$ if the system is beyond the boundary. In this way, the force on the system for $s < s_w$ comes from metadynamics and molecular interactions. For $s > s_w$, it comes from the restraining potential and the molecular interactions. Using this procedure, the history-dependent potential $V_G$ reaches a stationary state, and its time average in Eq. 9 becomes meaningful. This average value converges to the negative of the free energy everywhere, except in a region close to $s_w$, of size of the order of the Gaussian width $\delta s$. This procedure is simple and efficient, but it works only if the CV is one dimensional.

Another procedure for correcting the systematic errors that is accurate even close to the boundary is described in Crespo et al. (2010). In this procedure, if the system is in $s$, close to a boundary located in $s_w$, one extra Gaussian is added outside the boundary, in a position $2s_w - s$, specular to $s$ with respect to the boundary. The scope of this extra Gaussian is iteratively imposing that, in a suitably interval around $s_w$,

$$V_G(s_w - s, t) \approx 2V_G(s_w, t) - V_G(s_w + s, t) \tag{12}$$

This property ensures that, at stationary conditions, the history-dependent potential is approximately linear close to the boundary, but it does not impose the value of its derivative, which is iteratively determined by the thermodynamic bias. In practice, this condition is implemented as follows. If $|s - s_w| < \chi$, the additional Gaussian has a height $w$. $\chi$ is a free parameter that is normally chosen equal to the Gaussian width. If $|s - s_w| > \chi$, the height of the extra Gaussian is given by

$$(V_G(s_w - s, t) - 2V_G(s_w, t) + V_G(s_w + s, t)) \, y(s_w - s) \tag{13}$$

where the function $y(x) = \left(1 + \left(\frac{s_w - s}{4\chi}\right)^{10}\right)^{-1}$ is approximately one for $x < 4\chi$ and goes to zero for $x > 4\chi$. This ensures that $V_G$ goes smoothly to zero in the unphysical region. The procedure can be straightforwardly generalized to multidimensional CVs, but it leads to small systematic error at points near the joint boundaries of multiple CVs (Michael and de Pablo 2013).

A third approach for correcting the systematic errors at the boundaries was introduced in Michael and de Pablo (2013) and is possibly the most general, since it works in multiple dimensions and in domains of any shape. Consider a metadynamics performed in a finite domain $\Omega$ in arbitrarily many dimensions. The idea is modifying the Gaussian form of the bias potential in such a way

that, when the dynamics is at convergence, adding a new bias potential does not alter on average the free energy estimator. When metadynamics is at convergence in the region $\Omega$, the probability of observing the system in a point $s \in \Omega$ is $P_{\mathrm{conv}}(s) = 1/V_{\Omega} = 1/\int_{\Omega} ds$. One possible form of bias potential satisfying this property is

$$\tilde{g}(s, s') = w \frac{g(s - s')V_{\Omega}}{\int_{\Omega} ds' g(s - s')} \tag{14}$$

where $g(s - s') = \exp\left(-\frac{(s-s')^2}{2\delta s^2}\right)$. Indeed, it is straightforward to prove that $\int_{\Omega} ds' P_{\mathrm{conv}}(s')\tilde{g}(s, s') = w$ for all the possible choices of $s$. This condition automatically ensures that the bias potential becomes asymptotically well behaved as soon as the dynamics becomes diffusive in $\Omega$.

## 4 Well-Tempered Metadynamics

In this section it is described an alternative approach for obtaining a reliable estimator of the free energy, called well-tempered metadynamics. The idea at the basis of this approach is reducing iteratively the height of the Gaussians $w$. We will show that this framework allows deriving an exact estimator of the free energy, valid also at finite simulation time, namely, while $w \neq 0$.

The algorithm discussed in the previous section is a truly non-equilibrium algorithm, where the bias potential acting on the system is modified at a finite rate. For this reason, it is not true that $F(s) + V_G(s, t) = C(t)$. As discussed above, the free energy can be empirically estimated from the time average of $V_G(s, t)$, defined in Eq. 9. An exact relationship between $V_G(s, t)$ and $F(s)$ can be obtained if the rate at which the bias potential is modified is suitable decreased as the simulation progresses. The first algorithm exploiting this idea is well-tempered metadynamics (Barducci et al. 2008). Here, the update rule for the potential is modified as follows:

$$V_G(S(x), t) = w \sum_{\substack{t' = \tau_G, 2\tau_G, \cdots \\ t' < t}} \exp\left(-\frac{V_G(S(x), t')}{k_B \Delta T}\right)$$

$$\exp\left(-\sum_{\alpha=1}^{d} \frac{(S_\alpha(x) - s_\alpha(t'))^2}{2\delta s_\alpha^2}\right) \tag{15}$$

where $\Delta T$ is an additional parameter that has the units of a temperature. Note that often instead of $\Delta T$, one sees the equivalent parameter $\gamma = T + \frac{\Delta T}{T}$ in literature where $T$ is the temperature of the simulation. $\gamma$ is called bias or biasing factor. Clearly, by setting $\Delta T = \infty$, one recovers non-tempered metadynamics. However,

choosing a finite value of $\Delta T$, the rate at which the bias potential grows decreases with time. It has been shown that with this choice the rate decreases proportionally to one over the simulation time and makes the bias converge exactly provided that the simulation is sufficiently long (Barducci et al. 2008; Dama et al. 2014). The choice of including a "tempering factor" $\exp\left(-\frac{V_G(S(x),t')}{k_B \Delta T}\right)$ that depends on the value of $S$ has an interesting consequence. In fact, it can be shown that the potential converges to

$$V_G(s,t) = -\frac{\Delta T}{T + \Delta T} F(s) + C(t) , \qquad (16)$$

that is an a priori tunable fraction of the original free energy profile. In the limit where $\Delta T$ is vanishingly small, plain (unbiased) molecular dynamics is recovered. For a finite $\Delta T$, the residual barriers will make the system sample a probability distribution proportional to $\exp\left(-\frac{F(s)}{k_B T}\right) \exp\left(\frac{\Delta T}{T+\Delta T}\frac{F(s)}{k_B T}\right) = \exp\left(-\frac{F(s)}{k_B(T+\Delta T)}\right)$, which corresponds to increasing the temperature of the sampled CV. Thus, the effect of well-tempered metadynamics is similar to that of other non-equilibrium methods (VandeVondele and Rothlisberger 2002; Rosso et al. 2002; Maragliano and Vanden-Eijnden 2006) but is obtained with a quasi-equilibrium procedure. The influence of the choice of $\Delta T$ on the sampled distribution can be seen in Fig. 3. In short, the consequences of the well-tempered algorithm are twofold:

- The bias potential converges, due to the one-over-time schedule.
- The bias potential does not tend to cancel completely barriers but rather to partially compensate them.

These two factors can also be decoupled, and it is possible to design an algorithm where the potential converges rigorously but also cancels exactly the underlying free energy landscape (see, e.g., the globally tempered algorithm discussed in Dama et al. 2014).

It is interesting to observe that the one-over-time schedule of well-tempered metadynamics is commonly used in stochastic optimization problems, since it allows minimization to converge without remaining stuck. Interestingly, in the stochastic optimization literature, a common recipe is to estimate the parameters from their time average during the optimization procedure. In the case of well-tempered metadynamics, this would amount in using the time average of the bias potential in order to estimate the free energy profile, as it is commonly done in non-tempered metadynamics (see above). Whereas this is not strictly necessary, it can be done in order to decrease the numerical noise of the estimator. We also notice that computing the free energy profiles from the bias potential and comparing results for different simulation lengths might be misleading. Similarly to a cumulative average, such an estimator will converge by construction. In order to estimate errors, reweighting procedures in combination with block analysis are more reliable (see Sect. 4.1).
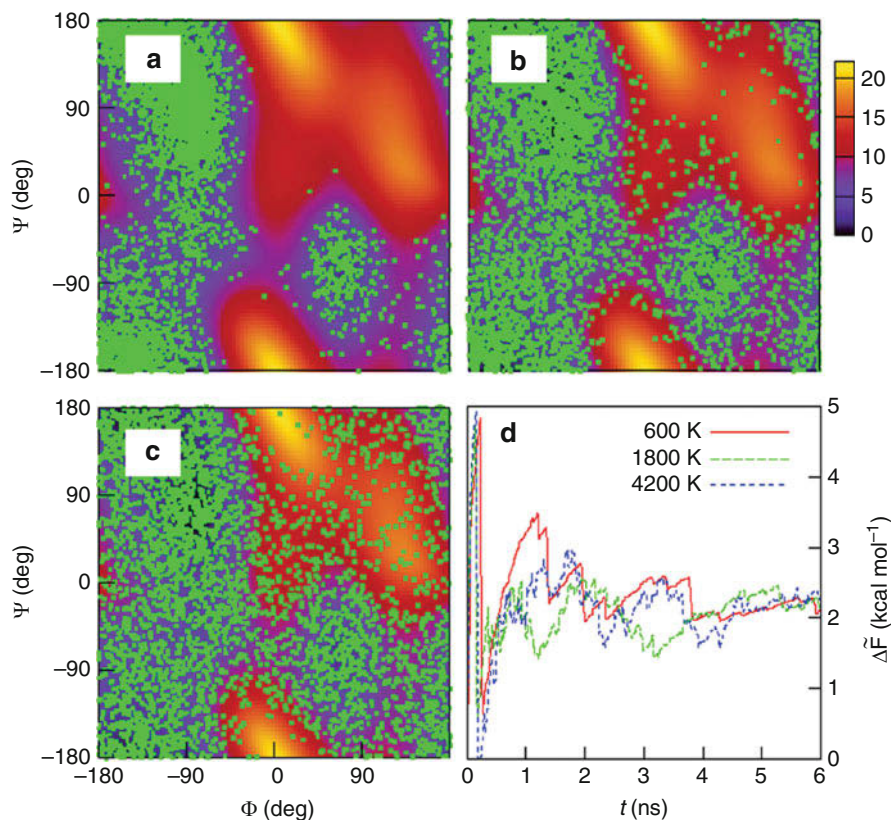
**Fig. 3** Panels (**a**–**c**): Representative trajectories obtained using well-tempered-metadynamics for different choices of $\Delta T$ (see text for definition) represented in the $(\phi, \psi)$ space for alanine dipeptide in vacuum. Green dots represent the visited conformations. Colors indicate the free energy profile. Panel (**d**): Estimate of the free energy difference between two metastable minima as a function of the simulation time. The estimate, obtained from the bias potential using Eq. 16, converges to the correct value. (After Barducci et al. 2008)

A simple interpretation of the fact that, in a well-tempered metadynamics simulation, the distribution of the CV at long times is not flat is the following. Since the prefactor for the accumulated Gaussians depends on the value of $S$, Gaussians of different heights are placed in different regions of the CV space. In order to reach a stationary distribution, it is thus necessary that the system spends more time in regions where small Gaussians are added and less time in regions where large Gaussians are added. This idea can be pushed further and used to convert metadynamics in an algorithm that is not designed to flatten completely (as in non-tempered metadynamics) or partially (as in well-tempered metadynamics) the histogram of the CVs but rather to enforce a predefined distribution (White et al. 2015; Marinelli and Faraldo-Gómez 2015). This application departs from the usual

scope of metadynamics simulations, which is enhancing sampling, and allows this technique to be used, for instance, to enforce experimental distributions (White et al. 2015; Marinelli and Faraldo-Gómez 2015; Gil-Ley et al. 2016). Notably, it can be shown that this approach is equivalent to a maximum entropy approach (Pitera and Chodera 2012) where the variables of which the average is constrained are the populations of each of the bins in the enforced distribution.

A further comment about the well-tempered formulation of metadynamics is that it is easy to see that the conventional estimator for the free energy Eq. 3 is, in the limit of Gaussians with a small width, completely equivalent to this umbrella sampling relation:

$$F(s) = -k_B T \log N(s) - V_G(s) \tag{17}$$

Here $N(s)$ is the histogram accumulated along the biased trajectory. This latter expression has however a fundamental advantage from the practical standpoint, namely, that it can be used also in the following cases (Branduardi et al. 2012):

- When the width of the Gaussians is too large to represent features of the free energy landscape.
- When one is willing to analyze a variable that is not equivalent to the one biased during the metadynamics simulation (reweighting), as it will be discussed in the next section.

These cases should be treated with some care. In the first case, if the free energy landscape contains significant barriers that can only be compensated by adding narrow Gaussians, clearly those barriers will not be sampled correctly. The second case is even more critical: if the analyzed (and not biased) variable displays significant barriers that were not compensated by a bias potential, its distribution might be very difficult to converge. A detailed analysis of finite width corrections to free energy estimation in well-tempered metadynamics was recently provided in Tiwary et al. (2015a).

## 4.1    Reweighting from Well-Tempered Metadynamics

The reweighting operation in metadynamics is one of its many useful features as it allows projecting probability densities on arbitrary CVs without having to repeat the simulation. As discussed in previous sections, in metadynamics one constructs a time-dependent bias $V(S(x), t)$ as a function of a low-dimensional CV $S(x)$. Here $x$ denotes the configurational coordinates. By assuming that the evolution of $V$ is very slow and that the system is thus at equilibrium, at time $t$ the biased probability distribution for $x$ is given by

$$P(x,t) = \frac{e^{-\beta[U(x)+V(S(x),t)]}}{\int dx \, e^{-\beta[U(x)+V(S(x),t)]}} = P_0(x) \, e^{-\beta[V(S(x),t)-c(t)]} \tag{18}$$

where $U(x)$ is the potential energy of the system (Tiwary and Parrinello 2014), $P_0(x)$ is the unbiased Boltzmann probability density, and the function $c(t)$ is defined as

$$c(t) = \frac{1}{\beta} \log \frac{\int ds e^{-\beta F(s)}}{\int ds e^{-\beta(F(s)+V(s,t))}} \tag{19}$$

As shown in Tiwary and Parrinello (2014) and Valsson et al. (2016), $c(t)$ can be calculated by substituting the running estimate of $F(s)$ (Tiwary and Parrinello 2014) into Eq. 19:

$$c(t) = \frac{1}{\beta} \log \frac{\int ds \, \exp\left\{\left[\frac{\gamma}{\gamma-1}\beta V(s,t)\right]\right\}}{\int ds \, \exp\left\{\left[\frac{1}{\gamma-1}\beta V(s,t)\right]\right\}}. \tag{20}$$

Here $\gamma$ is the bias factor introduced previously. Using Eq. 20 in Eq. 18 , we can then easily calculate the distribution of any generic observable $O(x)$ over the unbiased ensemble from the metadynamics trajectory through

$$\langle O(x)\rangle_0 = \left\langle O(x)e^{\beta[V(S(x),t)-c(t)]} \right\rangle. \tag{21}$$

It is also possible to derive a similar relation that is valid in the case where only the orthogonal variables (not biased) are sampled at equilibrium, provided that an independent estimate of the free energy as a function of the biased CVs $\tilde{P}_0(s)$ is provided. Namely, one can construct an estimator of the unbiased probability $\tilde{P}(x)$ as

$$\tilde{P}(x) = \frac{\tilde{P}_0(S(x))}{\int dx' \, \delta(S(x') - S(x))} P(x) \tag{22}$$

It is interesting that this reweighted distribution is, among all the possible distributions, the one that is closest to the distribution observed during the simulation and at the same time results in a marginal probability as a function of $s$ that is identical to $\tilde{P}_0(s)$, in a maximum entropy spirit (Pitera and Chodera 2012; White et al. 2015; Marinelli and Faraldo-Gómez 2015). By replacing the free energy estimator in Eq. 16, one obtains $\tilde{P}_0(s) \propto e^{-\frac{F(s)}{k_B T}} \propto e^{\frac{V_G(s)}{k_B T}\frac{T+\Delta T}{\Delta T}}$ where $V_G(s)$ is the final bias potential. Similarly, it is possible to notice that the denominator in Eq. 22 is the histogram of the CV as observed during the metadynamics simulations, which, in well-tempered metadynamics, is proportional to $e^{-\frac{V(s)}{k_B \Delta T}}$ (Barducci et al. 2008). By straightforward algebra, one can obtain the following estimator:

$$\tilde{P}(x) \propto e^{\frac{V(S(x))}{k_B T}} P(x) \tag{23}$$

Averages of arbitrary variables can be computed as

$$\langle O(x) \rangle_0 = \frac{\left\langle O(x) e^{\beta[V_G(S(x))]} \right\rangle}{\left\langle e^{\beta[V_G(S(x))]} \right\rangle} \tag{24}$$

where the bias potential at the end of the trajectory should be considered. This scheme has been proposed in Branduardi et al. (2012) to analyze simulations performed with Gaussian potentials larger than the typical features of the simulated landscape but can be used to reweight any CV.

A third option for reweighting has been introduced in an earlier work Bonomi et al. (2009a), although it is less practical than the two methods discussed above since it requires accumulating joint histograms for the biased and non-biased CVs.

It is interesting to compare the two alternative schemes presented in Eqs. 21 and 24. In the latter, the weight of each snapshot only depends on the value of the biased CVs for that specific snapshot, whereas in the former it depends on the specific form of the bias potential at the time when the snapshot was collected. This means that Eq. 24 is practical when analyzing a posteriori a trajectory, since one is not required to synchronize the history-dependent potential with the saved trajectory. However, Eq. 21 is the only option for on-the-fly analysis. The two procedures are expected to provide equivalent results in the long time limits, but performances of the two methods might be different if the simulation is not fully converged. This connects to a powerful yet not very widely used aspect of metadynamics and its self-diagnostic features. There are different schemes to obtain estimates of the converged free energy from a metadynamics run. While agreement in the estimates made using these different schemes is not a sufficient condition to judge the convergence of metadynamics, it clearly is a necessary condition. We thus encourage the user to make use of these different estimators – which are all implemented in a seamless and easy-to-use manner in PLUMED (Tribello et al. 2014) – for judging the quality and reliability of the free energy estimate made in metadynamics.

An important property of reweighting methods is that they basically allow free energies to be computed directly from histograms. These histograms can be compared across as-independent-as-possible fractions of the simulated trajectory (e.g., first half vs second half) or, to rigorously quantify the statistical error, using a block analysis. Clearly, results from different blocks will be consistent only if the same metastable states have been repeatedly visited.

## 4.2    Kinetic Rate Constants from Well-Tempered Metadynamics

One of the recent and very powerful developments in well-tempered metadynamics has been the demonstration of how to obtain unbiased kinetic properties from biased

metadynamics simulations of processes with very high barriers. Here we summarize the key ideas in this approach. We mention its requirements for applicability and how one may ascertain a posteriori whether or not these requirements have been met. We then mention some recent powerful applications of this approach to obtain timescales for difficult rare event problems such as drug unbinding and liquid droplet nucleation.

The method to obtain unbiased timescales from suitably biased metadynamics simulations has been called "infrequent metadynamics" (Tiwary and Parrinello 2013; Salvalaglio et al. 2014; Tiwary and Berne 2016b), as it involves periodic but infrequent biasing of a CV in order to increase the probability of escape from metastable states where the system would ordinarily be trapped for extended periods of time. If the biased CV can demarcate all relevant stable states of interest and if the time interval between biasing events is much longer compared to the transition path time spent in the ephemeral transition state (TS) regions, then one increases the probability of not adding bias in the TS regions and thereby not corrupting the dynamics during barrier crossing. This in turn preserves the sequence of transitions between stable states that the unbiased trajectory would have taken. The acceleration $\alpha$ of transition rates through biasing which directly yields the true unbiased rates can be obtained directly (i.e., without having to converge the free energy) through the use of the following expression which was first proposed by Grubmueller in the context of conformational flooding and independently by Voter in the context of hyperdynamics (Voter 1997; Grubmüller 1995):

$$\alpha = \langle e^{\beta V(s,t)} \rangle_t \tag{25}$$

Here $s$ is the CV being biased, $V(s,t)$ is the bias experienced at time $t$, and the subscript $t$ indicates averaging under the time-dependent potential.

Clearly such an approach will work only if (a) there is a clean timescale separation between the time spent in the TS and in the basins and (b) the chosen CV does actually demarcate between all stable states of interest. One can verify a posteriori if these requirements were met by checking if the cumulative distribution function for the transition times obtained by the use of a time-dependent biasing potential corrected using Eq. 25 indeed conforms to a time-homogeneous Poisson distribution (Salvalaglio et al. 2014). By the use of methods such as SGOOP (Tiwary and Berne 2016c, 2017; Tiwary 2017), it has been shown how one can construct a CV which is likely to meet these requirements on the basis of preliminary metadynamics runs using a sub-optimal CV.

## 5    The Choice of CVs

Similar to other methods that reconstruct the free energy in a set of generalized coordinates, the reliability of metadynamics is strongly influenced by the choice of the CVs. Ideally the CVs should satisfy three properties:

- They should clearly distinguish between the initial state, the final state, and the intermediates.
- They should describe all the slow events that are relevant to the process of interest.
- Their number should not be too large; otherwise it will take a very long time to fill the free energy surface.

Clearly, the second and third condition might be mutually exclusive, and in many cases it can be very difficult to find a "good" set of CVs.

Biasing explicitly all "slow" degrees of freedom is important for (a) the speed at which metadynamics converges to a reliable free energy and (b) the possibility to reconstruct unbiased rate constants from metadynamics. As per the proof of Dama et al. (2014), one may think that in principle metadynamics is agnostic to the precise choice of CVs being used for sampling. That is, even with sub-optimal choices of CVs which have some overlap with all slow degrees of freedom, metadynamics will converge to the right free energy. However in practice it is extremely useful to improve the choice of CVs before attempting to converge a metadynamics simulation. A simple metadynamics run on an idealized model can be enlightening. Consider the Z-shaped two-dimensional free energy depicted in Fig. 4. If a metadynamics simulation is performed biasing only CV1 and neglecting CV2, the simulation, which is started in basin B, is not able to perform in due time a transition toward A, and metadynamics goes on overfilling this minimum. A transition is finally observed only when the height of the accumulated Gaussians will largely exceed the true barrier height. This behavior will continue indefinitely without ever reaching a situation in which the free energy grows evenly like in the example of Fig. 1. A similar behavior is often observed in practical applications of metadynamics to real-world problems and is a strong indication that an important CV is missing. In general, one can check a posteriori if the description provided by the chosen set of CVs is accurate, for instance, by using transition path sampling techniques (Dellago et al. 1998, 2002) or by performing a committor test (see Dellago et al. 1998, 2002) or, when using metadynamics to obtain kinetic properties, by performing a Kolmogorov-Smirnov test as proposed in Salvalaglio et al. (2014). Even without performing a specific check, an hysteretic behavior in the free energy reconstruction always signals the lack of a relevant CV. If, instead, the free energy grows "smoothly," it is likely that the set of variables is adequate.

From the above discussion, it should be clear that it is not easy to a priori identify the correct set of CVs for a given problem at hand, and in many cases, it is necessary to proceed by trial and error, attempting several metadynamics simulations with different combinations of variables. That said, a large number of methods drawing from diverse inspirations have been proposed over the years that aim to systematically identify a good set of CVs for any given problem. Given the wide range of different ideas that have been used to design these methods, it would take another chapter to discuss them in a complete manner. As such, here we just point out some of the references that the interested reader is encouraged to look
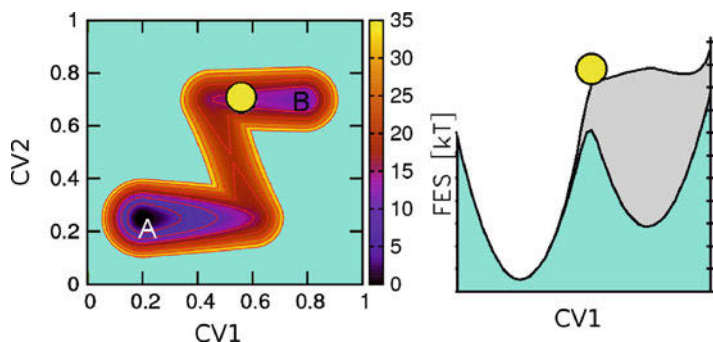
**Fig. 4** The effect of neglecting a relevant degree of freedom. Left side: 2D Z-shaped potential. Right side: the projection of the free energy on $s_1$ and a typical metadynamics bias generated using only $s_1$ as CV. Transitions from A to B are not described by $s_1$ causing strong hysteresis in the reconstructed free energy. (After Laio and Gervasio 2008)

up (Tribello et al. 2012; Rohrdanz et al. 2013; Tiwary and Berne 2016c, 2017; Sultan and Pande 2017; McCarty and Parrinello 2017).

# 6   Multi-replica Metadynamics Simulations

In this section, we discuss methods that can be used to combine metadynamics simulations performed simultaneously on separate replicas.

## 6.1   Multiple-Walkers Algorithms

The history-dependent potential used in metadynamics substantially depends on the histogram of the visited conformations in the CV space. The simulation can be straightforwardly parallelized by accumulating this potential *in parallel*. In other words, many simultaneous simulations could be used to update the bias potential that is then felt by all of them. The first application of this principle is the so-called multiple-walkers algorithm (Raiteri et al. 2006). In the non-tempered formulation of metadynamics, it has been shown that the error in the estimator of the free energy profile is independent of the number of walkers (Raiteri et al. 2006; Bussi et al. 2006b). Thus, a large number of walkers can be used to speed up the filling of the potential without affecting the final accuracy of the result. Walkers should only communicate when a new Gaussian is added to the bias potential (i.e., typically every few hundred steps), so that communication does not usually introduce any significant overhead. However, one should make sure that the walkers are initialized as independently as possible from each other. Similar properties are found when using well-tempered metadynamics (Barducci et al. 2008).

Interestingly, this idea has been recently extended to the case where series of similar (though not identical) systems are studied. For instance, if the free energy landscape associated to a protein-ligand binding should be evaluated for a number of different ligands that are similar to each other, then simulations can be performed simultaneously and the bias potential constructed by using also the information obtained with the other ligands. This scheme has been called altruistic metadynamics (Hosek et al. 2016) and requires the definition of an additional parameter, the altruistic factor, which weights the contribution of different simulations to a given bias potential. In its original formulation, the accuracy of the result is slightly deteriorated for large values of this altruistic factor. However, we suggest that by using reweighting schemes where the strict relationship between bias potential and free energy landscape is not required (as, e.g., in Branduardi et al. 2012) systematic errors might be decreased or eliminated.

## 6.2   Parallel Tempering Metadynamics

Parallel tempering metadynamics (PTmetaD) (Bussi et al. 2006a) is another approach based on running several metadynamics in parallel on different replicas. Multiple replicas of metadynamics are run at different temperature. An exchange of the coordinates of two replicas at adjacent temperatures is attempted with a fixed frequency, like in ordinary parallel tempering (Merlitz and Wenzel 2002). The acceptance ratio for an exchange involving replicas $i$ and $j$ is

$$P = \min\left\{1, \exp\left[(\beta_j - \beta_i)(V(x_j) - V(x_i))\right.\right.$$
$$\left.\left. + \beta_i(V_{G_i}(s(x_i)) - V_{G_i}(s(x_j))) + \beta_j(V_{G_j}(s(x_j)) - V_{G_j}(s(x_i)))\right]\right\}$$

where $V$ is the potential energy, $x_i$ are the coordinates of replica $i$, $\beta_i$ is the inverse temperature of replica $i$, and $V_{G_i}$ is the metadynamics potential of replica $i$ before the exchange. If the move is accepted, the coordinates are exchanged.

In this approach, the free energy profile as a function of the same collective variable $s$ is filled in parallel at all temperatures. Combining parallel tempering and metadynamics in this way improves the convergence of the free energy even when the CV is nonoptimal. Indeed, parallel tempering improves the sampling over the degrees of freedom not explicitly included in the CV space. At the same time, the approach converges more quickly than ordinary parallel tempering, since metadynamics enforces the exploration of high free energy configurations, ultimately leading to a more reliable estimate of the height of the free energy barriers.

A similar scheme can be used to combine solute tempering (Camilloni et al. 2008) or other variants of parallel tempering with metadynamics.

## 6.3    Bias Exchange

One of the reasons that make the choice of the CVs in metadynamics delicate is that they cannot be more than two or three in practical applications. Indeed, the time required to escape from a local minimum in the free energy surface is determined by the number of Gaussians that are necessary to fill the well. This number is proportional to $(1/\delta s)^d$, where d is the number of CVs that are biased. Hence, the efficiency of the method decreases exponentially with the number of dimensions involved. This applies both to ordinary metadynamics and to its well-tempered version.

A possible manner of addressing this problem is offered by bias exchange (Piana and Laio 2007) (BE), another technique based on the combined use of replica exchange and metadynamics. In this approach, one runs in parallel, at the same temperature, a large number $NR$ of metadynamics, with a bias potential acting on different CVs. The bias in each replica typically acts only on a single CV. Therefore the free energy wells can be filled efficiently. If all the variables were relevant for describing the process and the replicas were run independently, all the metadynamics would be affected by hysteresis, like in the example in Fig. 4. In bias exchange, at fixed time intervals, one attempts to swap the bias potentials between pairs of replicas. Let us denote by $s_k$, $k = 1, \cdots, NR$ the NR different CVs biased on the NR replicas. For two replicas of coordinates $x_i$ and $x_j$ and bias potentials $V_G(s_k(x_i), t)$ and $V_G(s_l(x_j), t)$, an exchange move consists in swapping $V_G(s_k, t)$ and $V_G(s_l, t)$. The move is accepted with a probability:

$$P = \min \left\{ 1, \exp \left[ \beta(V_G(s_k(x_i), t) + V_G(s_l(x_j), t) \right.\right.$$

$$\left.\left. - V_G(s_k(x_j), t) - V_G(s_l(x_i), t)) \right] \right\}$$

In this way, each replica is sequentially biased by a bias potential acting on one CV at a time. When a swap is accepted, the bias potential acting on two replicas changes direction. Simultaneously, the values of the CVs involved in the swap perform a jump. If, for example, an exchange between replica 1 and 2 is accepted, the CV biased on replica 1 will jump from $s_1(x_1)$ to $s_1(x_2)$, where $x_1$ and $x_2$ are the coordinates of the atoms in replica 1 and 2 before the exchange. These jumps greatly help decorrelating the dynamics and have the effect of improving the accuracy of the free energy estimate, greatly reducing the hysteresis. Of course, hysteresis can be eliminated only if all the relevant variables are included and are biased by at least one replica. Thus bias exchange does not eliminate the necessity of selecting before the simulation an appropriate set of CVs. Still, the approach allows treating simultaneously much more CVs than normal metadynamics. Therefore, it is of great help in setting up a simulation assuming as little as possible on the reaction mechanism.

It should also be remarked that in BE the bias potential grows in order to compensate NR one-dimensional projections of the free energy. In order to obtain multidimensional projections, it is necessary to combine data from all the replicas. This can be done using the WHAM method (Kumar et al. 1995) as explained in detail in Marinelli et al. (2009). A graphic user interface has been developed specifically to perform this task (Biarnés et al. 2012). In a typical BE simulation performed with a few tens of replicas, one might be able to reconstruct the free energy as a function of a few selected CVs. Reconstructing the free energy as a function of all the biased CVs is usually not possible (since it would require accumulating a histogram on a very high-dimensional space). In addition, although it is possible to compute free energies with respect to a posteriori chosen non-biased CVs, one should be aware that some of them might be hindered by non-flattened barriers and, as such, might not be sampled correctly.

The procedure discussed above to combine results from the multiple metadynamics simulations used in BE can be used to combine arbitrarily biased simulations. For instance, a bias-exchange-like procedure was used in Cunha and Bussi (2017) where each replica was running a metadynamics simulations biasing the distance between a ion and a possible binding site. However, in each replica, the coordination of the ion with the other binding sites was hindered by a repulsive potential. This straightforward generalization shows that arbitrary potentials can be added guided by the knowledge of the system and by the results of preliminary calculations and later reweighted using the WHAM method.

## 6.4    Other Ways to Bias a Large Number of CVs

Bias-exchange metadynamics is a commonly used algorithm that allows a large number of CVs to be biased and has been used already in a large number of applications. We here discuss two alternative approaches that emerged more recently.

The first approach is *parallel-bias* metadynamics (Pfaendtner and Bonomi 2015). This approach can be seen as a single-replica variant of the bias-exchange metadynamics protocol. Here, a possibly large number of CVs are defined that are alternatively biased (one at a time) during the simulation. An integer index keeps track of which is the currently biased CV and is evolved using a MC procedure. This procedure resembles the simulated-tempering scheme (Marinari and Parisi 1992), where a single simulation is performed and the temperature of the system is allowed to change in order to enhance sampling. In simulated-tempering simulations, an iterative procedure is required in order to force the simulation to spend the same amount of time at each temperature (see, e.g., Park and Pande 2007). On the contrary, in parallel-bias metadynamics, this effect is an implicit consequence of the self-healing character of the metadynamics bias potential. Indeed, if the simulation spends a large fraction of time biasing one specific CV, the bias potential acting on that CV will grow and will thus favor a transition such that a different CV will be biased. It should be said that in its practical implementation, the extra variable

corresponding to the index of the biased CV is usually marginalized so that it is not necessary to track it. As a consequence, the bias potential acting on the system reads

$$V_{PB}(s, t) = -\frac{1}{\beta} \log \sum_{\alpha} \left( -\beta V_G^{(\alpha)}(s_\alpha, t) \right) \qquad (26)$$

where $V_G^{(\alpha)}$ is the bias potential accumulated based on the history of variable $s_\alpha$. The parallel-bias metadynamics approach can be used in a multiple-walkers fashion (see Sect. 6.1) in order to exploit a parallel machine. The advantage with respect to traditional bias-exchange metadynamics is that the number of required replicas becomes effectively independent of the number of biased CVs.

An alternative procedure has been called *replica exchange with CV tempering* (Gil-Ley and Bussi 2015). Here, multiple CVs are biased concurrently in each simulation. More precisely, bias potentials aimed at compensating barriers on individual CVs are constructed as if these barriers were independent of each other. Since concurrently biasing many variables leads to sampling a conformational space that is very different from the original Boltzmann distribution, unbiased results are recovered by constructing a ladder of replicas with progressively increasing values of the bias factor (see Sect. 4). The resulting bias potential on replica $i$ is thus

$$V_{RECT}^{(i)}(s, t) = \sum_{\alpha} \left( V_G^{(\gamma_i, \alpha)}(s_\alpha, t) \right) \qquad (27)$$

where $V_G^{(\gamma_i, \alpha)}$ is a potential accumulated based on the history of variable $s_\alpha$ and using a bias factor $\gamma_i$. The $\gamma_i$ corresponding to the lowest replica should be set equal to 1 in order to allow unbiased results to be recovered, whereas the one corresponding to the highest replica should be large enough for the important energy barriers to be crossed. Similarly to other replica exchange-based methods, the acceptance rate should take into account the different bias potentials acting on different replicas. Importantly, the accuracy of the final result is independent of the assumption that CVs are independent of each other. We notice that, at variance with parallel-bias metadynamics and bias-exchange metadynamics, replica exchange with CV tempering concurrently accelerates all the chosen CVs. At variance with bias-exchange metadynamics, the number of required replicas scales with the square root of the number of CVs. However, concurrently biasing too many CVs could also be counterproductive since it might enlarge too much the extension of the conformational space that should be sampled in order to converge a simulation. On the other hand, it might be more effective when biasing simple variables that are not well correlated with the overall barriers hindering the free energy landscape.

# 7    Discussion and Outlook

Metadynamics was introduced in 2002 as a method to escape and sample free energy minima in complex molecular systems. Over the years, it has grown significantly in its variants and in range of applications and have been implemented in several software plugins and analysis tools (Bonomi et al. 2009b; Henin et al. 2009; Fiorin et al. 2013; Giorgino et al. 2017; Sidky et al. 2018). In this book chapter, we have attempted to summarize the state of the art concerning the different forms and formulations of metadynamics, with a focus on the theory rather than the applications. We did not discuss all problems in full detail – for example, the methods for automatic construction of the CVs – an important problem at the center of several efforts in the community. In this final section, we conclude with our outlook on some *do*s and *don′t*s concerning metadynamics in its traditional use for estimating free energy profiles, drawing from the experience of the three authors. Our discussion mainly concerns two vital points regarding any metadynamics simulation which we have already discussed in this chapter and would like to emphasize here: (a) the choice of the appropriate variant of metadynamics according to the dimensionality of the CV space and (b) ascertaining when the sampled distribution has converged and the simulation is meaningful. The discussion in this chapter is clearly limited for space reason. Interested readers can refer to the many available reviews dedicated to the metadynamics method (Laio and Gervasio 2008; Barducci et al. 2011; Bussi et al. 2015; Valsson et al. 2016; Tiwary and van de Walle 2016).

- In general, well-tempered metadynamics is an apt choice if the number of CVs is less than or equal to 3. This approach requires choosing one extra parameter with respect to ordinary metadynamics (the CV temperature $\Delta T$) that allows a tuning of the explored region, and its convergence has been proved rigorously (Barducci et al. 2008; Dama et al. 2014). When using more than 3 CVs, bias exchange or other methods based on the idea of reconstructing multiple low-dimensional potentials should be preferred. Indeed, accumulating statistics in a high-dimensional space to construct a bias potential capable to enhance sampling could be virtually impossible. In cases when one-dimensional projections of the free energy are barrierless, non-tempered metadynamics should be preferred, since the quasi-static bias potential obtained at the end of a well-tempered metadynamics simulation would likely not help in enhancing transitions and the simulation could remain stuck. On the contrary, non-tempered metadynamics would effectively increase the diffusion coefficient in the CV space by forcing it to sample equivalently all the points.
- Standard metadynamics should always be used together with the appropriate boundary conditions on the CV, as described in Sect. 3.1. The free energy should be always estimated as a time average of the history-dependent potential, controlling the accuracy by block analysis.

- In metadynamics simulations (as much as in all molecular dynamics simulations), it is very important to check if transitions between "reactants" and "products" are actually observed during the simulation. In case a very small number of such transitions are detected, one should carefully assess the convergence and the reproducibility of the results. Special attention should be paid in analyzing replica exchange simulations, where continuous trajectories should be analyzed (in opposition to the discontinuous trajectories corresponding to each replica).
- When analyzing well-tempered-metadynamics, it is important to recognize that the changes in the bias potential are decreasing by construction since the Gaussian height decreases as one over time. Computing the free energy profile from the bias potential and comparing results for different simulation length might be misleading if there are not multiple transitions observed as discussed above. Performing a block analysis on a reweighting procedure can provide more robust error estimations (see Sect. 4.1).

While in this chapter we did not discuss in detail the various applications of metadynamics to materials modeling problems, here we would like to point out that the metadynamics approach has been extensively used from the very beginning for modeling materials and predicting their properties. For example, the approach was used to characterize the properties of bulk materials, in their crystalline and amorphous phases (Donadio et al. 2005; Donadio and Bernasconi 2005; Di Pietro et al. 2006; Behler et al. 2008; Hu et al. 2012; Pietrucci et al. 2010), and to study nucleation (Trudu et al. 2006; Quigley and Rodger 2008; Piaggi et al. 2017; Salvalaglio et al. 2015, 2016; Fitzner et al. 2017; Mendels et al. 2018), charge diffusion (Laino et al. 2007; Iannuzzi and Parrinello 2004; Vartak et al. 2013), interaction of small molecules with surfaces or porous materials (Palafox-Hernandez et al. 2014; Hasell et al. 2016; Boyer et al. 2016; Munro et al. 2016; Bui et al. 2017), role of solvent in nanoassembly (Tiwary et al. 2015b; Tiwary and Berne 2016a), and hydrogen fuel cells (Cheng et al. 2017), just to name a few. Another field related to material science in which the approach has been applied with great success is crystal structure prediction (Martoňák et al. 2003, 2005, 2006, 2007; Ceriani et al. 2004; Oganov et al. 2005; Karamertzanis et al. 2008). Finally, the method can be used for studying chemical reactions on surfaces and heterogeneous catalysis (Ensing et al. 2006; Zipoli et al. 2004; Fleming et al. 2016; Oliveira et al. 2018; Fu and Pfaendtner 2018).

In conclusion, metadynamics has now become a standard tool in the workbox of a practitioner interested in sampling molecular systems. Many of the original intuitive formulations have now been demonstrated on a sound theoretical footing, and many variants have also come up which have truly extended the applicability of the approach. The role of intuition – especially in the choice of low-dimensional CVs – has not yet been eliminated and possibly will never be completely gone. A very important property of metadynamics is that, if a simulation is analyzed correctly, it is relatively easy to identify errors due to incorrect sampling. In many cases, problematic simulations can be used in order to choose better CVs for a next

round of simulations. This is not true in the case of the standard umbrella sampling protocol performed combining multiple independent restrained simulations, where hysteresis effects are often very difficult to detect. It has also become clear now that the infrequent metadynamics formulation can safely and self-consistently be used to compute unbiased kinetic rate constants in a wide class of rare event systems. We thus end on a note of optimism for the future use of metadynamics in tackling diverse and relevant problems in science and technology.

# References

Baftizadeh F, Cossio P, Pietrucci F, Laio A (2012) Protein folding and ligand-enzyme binding from bias-exchange meta-dynamics simulations. Curr Phys Chem 2:79–91

Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett 1(2):020603. https://doi.org/10.1103/PhysRevLett.100.020603

Barducci A, Bonomi M, Parrinello M (2011) Metadynamics. Wiley Interdiscip Rev Comput Mol Sci 1(5):826–843

Behler J, Martonak R, Donadio D, Parrinello M (2008) Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. Phys Rev Lett 100(18):185501. https://doi.org/10.1103/PhysRevLett.100.185501

Biarnés X, Pietrucci F, Marinelli F, Laio A (2012) METAGUI. A VMD interface for analyzing metadynamics and molecular dynamics simulations. Comput Phys Commun 183(1):203–211. https://doi.org/10.1016/j.cpc.2011.08.020

Bonomi M, Barducci A, Parrinello M (2009a) Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. J Comput Chem 30(11):1615–1621

Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA et al (2009b) Plumed: a portable plugin for free-energy calculations with molecular dynamics. Comput Phys Commun 180(10):1961–1972

Boyer MJ, Vilčiauskas L, Hwang GS (2016) Structure and li+ ion transport in a mixed carbonate/lipf 6 electrolyte near graphite electrode surfaces: a molecular dynamics study. Phys Chem Chem Phys 18(40):27868–27876

Branduardi D, Bussi G, Parrinello M (2012) Metadynamics with adaptive gaussians. J Chem Theory Comput 8(7):2247–2254

Bui T, Phan A, Cole DR, Striolo A (2017) Transport mechanism of guest methane in water-filled nanopores. J Phys Chem C 121(29):15675–15686

Bussi G, Gervasio FL, Laio A, Parrinello M (2006a) Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. J Am Chem Soc 128(41):13435–13441. https://doi.org/10.1021/ja062463w

Bussi G, Laio A, Parrinello M (2006b) Equilibrium free energies from nonequilibrium metadynamics. Phys Rev Lett 96(9):090601. https://doi.org/10.1103/PhysRevLett.96.090601

Bussi G, Branduardi D et al (2015) Free-energy calculations with metadynamics: theory and practice. Rev Comput Chem 28:1–49

Camilloni C, Provasi D, Tiana G, Broglia RA (2008) Exploring the protein G helix free-energy surface by solute tempering metadynamics. Proteins Struct Funct Bioinf 71(4):1647–1654. https://doi.org/10.1002/prot.21852

Car R, Parrinello M (1985) Unified approach for molecular-dynamics and density-functional theory. Phys Rev Lett 45:2471

Carter EA, Ciccotti G, Hynes JT, Kapral R (1989) Constrained reaction coordinate dynamics for the simulation of rare events. Chem Phys Lett 156:472–477

Ceriani C, Laio A, Fois E, Gamba A, Martonak R, Parrinello M (2004) Molecular dynamics simulation of reconstructive phase transitions on an anhydrous zeolite. Phys Rev B 70(11):113403. https://doi.org/10.1103/PhysRevB.70.113403

Cheng T, Goddard WA, An Q, Xiao H, Merinov B, Morozov S (2017) Mechanism and kinetics of the electrocatalytic reaction responsible for the high cost of hydrogen fuel cells. Phys Chem Chem Phys 19(4):2666–2673

Crespo Y, Marinelli F, Pietrucci F, Laio A (2010) Metadynamics convergence law in a multidimensional system. Phys Rev E 81:055701. https://doi.org/10.1103/PhysRevE.81.055701

Cunha RA, Bussi G (2017) Unraveling $Mg^{2+}$–Rna binding with atomistic molecular dynamics. RNA 23(5):628–638

Cvijovic D, Klinowski J (1995) Taboo search – an approach to the multiple minima problem. Science 267:664–666

Dama JF, Parrinello M, Voth GA (2014) Well-tempered metadynamics converges asymptotically. Phys Rev Lett 112(24):240602

Dellago C, Bolhuis P, Csajka FS, Chandler D (1998) Transition path sampling and the calculation of rate constants. J Chem Phys 108:1964–1977

Dellago C, Bolhuis P, Geissler P (2002) Transition path sampling. Adv Chem Phys 123:1–78

Di Pietro E, Pagliai M, Cardini G, Schettino V (2006) Solid-state phase transition induced by pressure in LiOH center dot $H_2O$. J Phys Chem B 110(27):13539–13546. https://doi.org/10.1021/jp061620a

Donadio D, Bernasconi M (2005) Ab initio simulation of photoinduced transformation of small rings in amorphous silica. Phys Rev B 71(7):073307. https://doi.org/10.1103/PhysRevB.71.073307

Donadio D, Raiteri P, Parrinello M (2005) Topological defects and bulk melting of hexagonal ice. J Phys Chem B 109:5421–5424

Ensing B, De Vivo M, Liu Z, Moore P, Klein ML (2006) Metadynamics as a tool for exploring free energy landscapes of chemical reactions. Acc Chem Res 39(2):73–81

Fiorin G, Klein ML, Hénin J (2013) Using collective variables to drive molecular dynamics simulations. Mol Phys 111(22–23):3345–3362

Fitzner M, Sosso GC, Pietrucci F, Pipolo S, Michaelides A (2017) Pre-critical fluctuations and what they disclose about heterogeneous crystal nucleation. Nat Commun 8(1):2257

Fleming KL, Tiwary P, Pfaendtner J (2016) New approach for investigating reaction dynamics and rates with ab initio calculations. J Phys Chem A 120(2):299–305

Fu CD, Pfaendtner J (2018) Lifting the curse of dimensionality on enhanced sampling of reaction networks with parallel bias metadynamics. J Chem Theory Comput 14:2516–2525

Gil-Ley A, Bussi G (2015) Enhanced conformational sampling using replica exchange with collective-variable tempering. J Chem Theory Comput 11(3):1077–1085

Gil-Ley A, Bottaro S, Bussi G (2016) Empirical corrections to the amber RNA force field with target metadynamics. J Chem Theory Comput 12(6):2790–2798

Giorgino T, Laio A, Rodriguez A (2017) METAGUI 3: a graphical user interface for choosing the collective variables in molecular dynamics simulations. Comput Phys Commun 217:204–209

Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: conformational flooding. Phys Rev E 52(3):2893

Hasell T, Miklitz M, Stephenson A, Little MA, Chong SY, Clowes R, Chen L, Holden D, Tribello GA, Jelfs KE et al (2016) Porous organic cages for sulfur hexafluoride separation. J Am Chem Soc 138(5):1653–1659

Henin J, Fiorin G, Chipot C, Klein ML (2009) Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. J Chem Theory Comput 6(1):35–47

Hosek P, Toulcova D, Bortolato A, Spiwok V (2016) Altruistic metadynamics: multisystem biased simulation. J Phys Chem B 120(9):2209–2215

Hu XL, Piccinin S, Laio A, Fabris S (2012) Atomistic structure of cobalt-phosphate nanoparticles for catalytic water oxidation. ACS Nano 6(12):10497

Huber T, Torda A, van Gunsteren W (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aided Mol Des 8:695–708

Iannuzzi M, Parrinello M (2004) Proton transfer in heterocycle crystals. Phys Rev Lett 93: 025901

Iannuzzi M, Laio A, Parrinello M (2003) Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. Phys Rev Lett 90:238302

Karamertzanis PG, Raiteri P, Parrinello M, Leslie M, Price SL (2008) The thermal stability of lattice-energy minima of 5-fluorouracil: metadynamics as an aid to polymorph prediction. J Phys Chem B 112(14):4298–4308. https://doi.org/10.1021/jp709764e

Kevrekidis IG, Gear CW, Hummer G (2004) Equation-free: the computer-aided analysis of comptex multiscale systems. AIChE J 50(7):1346–1355

Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1995) Multidimensional free-energy calculations using the weighted histogram analysis method. J Comput Chem 16:1339–1350

Laino T, Donadio D, Kuo IFW (2007) Migration of positively charged defects in alpha-quartz. Phys Rev B 76(19):195210. https://doi.org/10.1103/PhysRevB.76.195210

Laio A, Gervasio FL (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys 71(12):126601

Laio A, Parrinello M (2002) Escaping free energy minima. Proc Natl Acad Sci USA 99:12562–12566

Laio A, Rodriguez-Fortea A, Gervasio FL, Ceccarelli M, Parrinello M (2005) Assessing the accuracy of metadynamics. J Phys Chem B 109:6714–6721

Maragliano L, Vanden-Eijnden E (2006) A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. Chem Phys Lett 426(1):168–175

Marinari E, Parisi G (1992) Simulated tempering: a new Monte Carlo scheme. EPL (Europhys Lett) 19(6):451

Marinelli F, Faraldo-Gómez JD (2015) Ensemble-biased metadynamics: a molecular simulation method to sample experimental distributions. Biophys J 108(12):2779–2782

Marinelli F, Pietrucci F, Laio A, Piana S (2009) A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. PLOS Comput Biol 5:1–18. https://doi.org/10.1371/journal.pcbi.1000452

Martoňák R, Laio A, Parrinello M (2003) Predicting crystal structures: the Parrinello-Rahman method revisited. Phys Rev Lett 90:75503

Martoňák R, Laio A, Bernasconi M, Ceriani C, Raiteri P, Parrinello M (2005) Simulation of structural phase transitions by metadynamics. Z Krist 220:489–498

Martoňák R, Donadio D, Oganov AR, Parrinello M (2006) Crystal structure transformations in $SiO_2$ from classical and ab initio metadynamics. Nat Mater 5(8):623–626. https://doi.org/10.1038/nmat1696

Martoňák R, Donadio D, Oganov AR, Parrinello M (2007) From four- to six-coordinated silica: transformation pathways from metadynamics. Phys Rev B 76(1):014120. https://doi.org/10.1103/PhysRevB.76.014120

McCarty J, Parrinello M (2017) A variational conformational dynamics approach to the selection of collective variables in metadynamics. J Chem Phys 147(20):204109. https://doi.org/10.1063/1.4998598

Mendels D, McCarty J, Piaggi PM, Parrinello M (2018) Searching for entropically stabilized phases: the case of silver iodide. J Phys Chem C 122(3):1786–1790

Merlitz H, Wenzel W (2002) Comparison of stochastic optimization methods for receptor-ligand docking. Chem Phys Lett 362:271–277

Michael M, de Pablo J (2013) A boundary correction algorithm for metadynamics in multiple dimensions. J Chem Phys 139:084102

Micheletti C, Laio A, Parrinello M (2004) Reconstructing the density of states by history-dependent metadynamics. Phys Rev Lett 92:170601

Munro CJ, Hughes ZE, Walsh TR, Knecht MR (2016) Peptide sequence effects control the single pot reduction, nucleation, and growth of au nanoparticles. J Phys Chem C 120(33): 18917–18924

Oganov A, Martonak R, Laio A, Raiteri P, Parrinello M (2005) Anisotropy of Earth's D'' layer and stacking faults in the MgSiO3 post-perovskite phase. Nature 438(7071):1142–1144. https://doi.org/10.1038/nature04439

Oliveira LF, Fu CD, Pfaendtner J (2018) Density functional tight-binding and infrequent metadynamics can capture entropic effects in intramolecular hydrogen transfer reactions. J Chem Phys 148(15):154101

Palafox-Hernandez JP, Tang Z, Hughes ZE, Li Y, Swihart MT, Prasad PN, Walsh TR, Knecht MR (2014) Comparative study of materials-binding peptide interactions with gold and silver surfaces and nanostructures: a thermodynamic basis for biological selectivity of inorganic materials. Chem Mater 26(17):4960–4969

Park S, Pande VS (2007) Choosing weights for simulated tempering. Phys Rev E 76(1):016703

Patey GN, Valleau JP (1975) Monte-Carlo method for obtaining interionic potential of mean force in ionic solution. J Chem Phys 63:2334–2339

Pfaendtner J, Bonomi M (2015) Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. J Chem Theory Comput 11(11):5062–5067

Piaggi PM, Valsson O, Parrinello M (2017) Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. Phys Rev Lett 119(1):015701

Piana S, Laio A (2007) A bias-exchange approach to protein folding. J Phys Chem B 111(17):4553–4559. https://doi.org/10.1021/jp0678731

Pietrucci F, Gerra G, Andreoni W (2010) CdTe surfaces: characterizing dynamical processes with first-principles metadynamics. Appl Phys Lett 97(14):141914

Pitera JW, Chodera JD (2012) On the use of experimental observations to bias simulated ensembles. J Chem Theory Comput 8(10):3445–3451

Quigley D, Rodger PM (2008) Metadynamics simulations of ice nucleation and growth. J Comput Phys 128(15):154518. https://doi.org/10.1063/1.2888999

Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. J Phys Chem B 110:3533–3539

Risken H (1989) The Fokker-Planck equation. Springer

Rohrdanz MA, Zheng W, Clementi C (2013) Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. Ann Rev Phys Chem 64:295–316

Rosso L, Mináry P, Zhu Z, Tuckerman ME (2002) On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. J Comput Phys 116(11):4389–4402

Salvalaglio M, Tiwary P, Parrinello M (2014) Assessing the reliability of the dynamics reconstructed from metadynamics. J Chem Theor Comput 10(4):1420–1425. https://doi.org/10.1021/ct500040r

Salvalaglio M, Perego C, Giberti F, Mazzotti M, Parrinello M (2015) Molecular-dynamics simulations of urea nucleation from aqueous solution. Proc Natl Acad Sci 112(1):E6–E14

Salvalaglio M, Tiwary P, Maggioni GM, Mazzotti M, Parrinello M (2016) Overcoming time scale and finite size limitations to compute nucleation rates from small scale well tempered metadynamics simulations. J Chem Phys 145(21):211925

Sidky H et al, (2018) SSAGES: Software Suite for Advanced General Ensemble Simulations. J Chem Phys 148:044104 https://doi.org/10.1063/1.5008853

Sultan MM, Pande VS (2017) Tica-metadynamics: accelerating metadynamics by using kinetically selected collective variables. J Chem Theory Comput 13(6):2440–2447. https://doi.org/10.1021/acs.jctc.7b00182, pMID:28383914

Theodoropoulos C, Qian Y, Kevrekidis IG (2000) Coarse stability and bifurcation analysis using time-steppers: a reaction-diffusion example. Proc Natl Acad Sci USA 97:9840–9843

Tiwary P (2017) Molecular determinants and bottlenecks in the dissociation dynamics of biotin-streptavidin. J Phys Chem B 121(48):10841–10849. https://doi.org/10.1021/acs.jpcb.7b09510

Tiwary P, Berne B (2016a) How wet should be the reaction coordinate for ligand unbinding? J Chem Phys 145(5):054113

Tiwary P, Berne BJ (2016b) Kramers turnover: from energy diffusion to spatial diffusion using metadynamics. J Chem Phys 144(13):134103–134106

Tiwary P, Berne BJ (2016c) Spectral gap optimization of order parameters for sampling complex molecular systems. Proc Natl Acad Sci 113(11):2839–2844. https://doi.org/10.1073/pnas.1600917113

Tiwary P, Berne BJ (2017) Predicting reaction coordinates in energy landscapes with diffusion anisotropy. J Chem Phys 147(15):152701

Tiwary P, Parrinello M (2013) From metadynamics to dynamics. Phys Rev Lett 111:230602–230606. https://doi.org/10.1103/PhysRevLett.111.230602

Tiwary P, Parrinello M (2014) A time-independent free energy estimator for metadynamics. J Phys Chem B 119(3):736–742

Tiwary P, van de Walle A (2016) A review of enhanced sampling approaches for accelerated molecular dynamics. In: Multiscale materials modeling for nanomechanics. Springer, New York pp 195–221. https://doi.org/10.1007/978-3-319-33480-6_6

Tiwary P, Dama JF, Parrinello M (2015a) A perturbative solution to metadynamics ordinary differential equation. J Chem Phys 143(23):234112

Tiwary P, Mondal J, Morrone JA, Berne B (2015b) Role of water and steric constraints in the kinetics of cavity–ligand unbinding. Proc Natl Acad Sci 112(39):12015–12019

Tribello GA, Ceriotti M, Parrinello M (2012) Using sketch-map coordinates to analyze and bias molecular dynamics simulations. Proc Natl Acad Sci 109(14):5196–5201

Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) Plumed 2: new feathers for an old bird. Comput Phys Commun 185(2):604–613

Trudu F, Donadio D, Parrinello M (2006) Freezing of a Lennard-Jones fluid: from nucleation to spinodal regime. Phys Rev Lett 97(10):105701. https://doi.org/10.1103/PhysRevLett.97.105701

Valsson O, Tiwary P, Parrinello M (2016) Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. Ann Rev Phys Chem 67(1):159–184

VandeVondele J, Rothlisberger U (2002) Canonical adiabatic free energy sampling (cafes): a novel method for the exploration of free energy surfaces. J Phys Chem B 106(1):203–208

Vartak S, Roudgar A, Golovnev A, Eikerling M (2013) Collective proton dynamics at highly charged interfaces studied by ab initio metadynamics. J Phys Chem B 117(2):583–588

Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett 78:3908–3911. https://doi.org/10.1103/PhysRevLett.78.3908

Wang F, Landau DP (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. Phys Rev Lett 86:2050

White AD, Dama JF, Voth GA (2015) Designing free energy surfaces that match experimental data with metadynamics. J Chem Theory Comput 11(6):2451–2460

Zipoli F, Bernasconi M, Martoňák R (2004) Constant pressure reactive molecular dynamics simulations of phase transitions under pressure: the graphite to diamond conversion revisited. Eur Phys J B 39:41–47

# Novel Enhanced Sampling Strategies for Transitions Between Ordered and Disordered Structures

# 28

## Fabio Pietrucci

## Contents

**Abstract**

At the atomic scale, condensed matter displays a fascinating variety of structural transformation processes. Examples include phase transitions between ordered and/or disordered structures (crystal to crystal, liquid to crystal, amorphous to crystal, etc.), isomerization of nanoclusters, chemical reactions, protein conformational changes, and many other phenomena. In all these cases, it is necessary to find suitable distance metrics and collective variables in order to analyze atomistic simulations of transformations as well as to accelerate them with enhanced sampling techniques, yielding mechanisms and free-energy landscapes. In this context, the present chapter illustrates approaches stemming from the idea of watching transformations of matter as modifications of the adjacency matrix formed by interatomic connections. The resulting tools have a general formulation and can therefore be applied to a range of different processes in physics, chemistry, and nanoscience.

F. Pietrucci (✉)
Muséum National d'Histoire Naturelle, UMR CNRS 7590, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, Sorbonne Université, Paris, France
e-mail: fabio.pietrucci@upmc.fr

# 1    Introduction

A major challenge of computational materials science is the accurate prediction of the transition mechanisms connecting together different structures of a same material. These structures, both in extended solids or liquids and in nanostructures, range from perfectly ordered to highly disordered and typically correspond to minima separated by barriers in a free-energy landscape as a function of suitable order parameters. Very often, it is much easier to generate in silico a large number of putative structures than to characterize the transformation pathways between them, and this asymmetry has important practical consequences.

A prominent example is the synthesis of new crystalline structures with desired properties. In recent years, computational structure prediction methods (Glass et al. 2006; Pickard and Needs 2011; Wang et al. 2012) have contributed to the rapid increase of new predicted phases of materials, of high interest for potential applications (see, e.g., Wilmer et al. 2012). However, at present, no general theoretical approach has been developed to guide experiments along the possible pathways connecting stable structures, so that synthesis remains a challenging endeavor, slowed down by expensive trial-and-error cycles and often dominated by kinetics rather than thermodynamics.

In principle, a powerful tool for investigating transformation mechanisms and the corresponding energetics is provided by atomistic molecular dynamics or Monte Carlo simulations combined with enhanced sampling techniques that accelerate barrier-crossing and allow the reconstruction of free-energy landscapes. Unfortunately, available approaches for phase transitions and other structural transformations lack transferability. In other words, the study of each particular system and process typically requires a time-consuming preliminary optimization of the computational setup, often involving trial-and-error and several repetitions of the simulation. In particular, since most enhanced sampling techniques demand to identify system-specific collective variables (CVs) to accelerate the exploration of pathways and reconstruction of free-energy landscapes, finding suitable CVs is often the crucial and most cumbersome step in the computational study (Pietrucci 2017). At the basic level, such CVs must be capable of distinguishing the initial and final states of the transformation, as well as intermediate structures: from this point of view, disordered systems can be more problematic to analyze than ordered ones, due to the lack of evident symmetries.

Both in experiments and in silico, it is unclear, in general, how to find a viable pathway reaching a particular poly(a)morph (or nanostructure) different from the starting one, possibly separated by several intermediate steps. To complicate things, metastable forms are very often involved in phase transitions (Schreiber et al. 2017), and sometimes their kinetic stability is very high. Thus, in order to recover the desired structure, one needs to manipulate the control parameters, e.g., pressure and temperature, often in a complex and delicate way that is not trivial to guess (Radha et al. 2015). A precise understanding of transition mechanisms and the corresponding kinetics is therefore the key to explain and control the behavior of matter. In this chapter, some recent conceptual and practical developments are

presented that aim to bridge the gap between structure prediction and synthetic routes prediction, illustrated by means of different examples ranging from silicon and carbon nanostructures to water poly(a)morphism. As mentioned at the end, this class of methodologies can be very effective also for the study of chemical reactions in solution.

## 2  Representing Matter Through the Adjacency Matrix

Let us consider the possible transformation processes occurring in condensed matter, adopting, for the latter, a broad definition including atomic aggregates like molecules, nanostructures, liquids, and solids. These diverse systems undergo a variety of transformations including, e.g., phase transitions, chemical reactions, or protein conformational changes. From the viewpoint of theory and simulation, an interesting question is whether we can find a unified way to describe in atomic detail the transition pathways of all such different phenomena. One immediately imagines, building upon the basic Cartesian coordinates of atoms, many possible mathematical representations: examples are sets of distances, cell parameters (for periodic systems), coordination numbers, symmetry indicators, dihedral angles, root mean square deviations, etc. However, the preceding examples are typically tailored on specific systems and transformations. Is it possible to find a general representation, able to accurately track the detailed dynamic evolution of a wide range of transformations?

A possible starting point is to consider a generic atomic structure as a network: atoms are the nodes, while links can be defined based on proximity. Likewise, structural transitions become network transformations. The advantages of this view are its generality, encompassing many processes in condensed matter, and the availability of the powerful graph theory toolbox. In particular, a graph of $N$ nodes can be represented by a $N \times N$ adjacency matrix (also called contact matrix): $a_{ij} = 1$ if nodes $i$ and $j$ are connected and zero otherwise. The matrix contains the detailed topology of the network. Clearly, the "connection" between atoms appears as a somehow arbitrary concept: it can be chosen to reflect the existence of a chemical bond (ionic, covalent, hydrogen-bond, etc.), based on energy and electronic structure, or it can be introduced as an arbitrary property of close-enough atoms. Both choices have merits and for its simplicity we will follow the second one (that often is compatible with the first). In practice, it is convenient to introduce a switching function $f(d_{ij})$ of the interatomic distance, decaying smoothly from one to zero in a chosen range of distances (an example is the Fermi-Dirac function). The range can be inferred by inspecting pair distribution functions, e.g., so that $f$ differentiates bonded atomic pairs from second-shell pairs. Note that replacing plain distances with a monotonic function of the distances preserves the same information, except when distances are so large (or so small) that the derivative of $f$ is close to zero.

The resulting adjacency matrix is composed by real-valued elements $a_{ij} \equiv f(d_{ij}) \in [0, 1]$: examples for different systems are shown in Fig. 1. In this way,
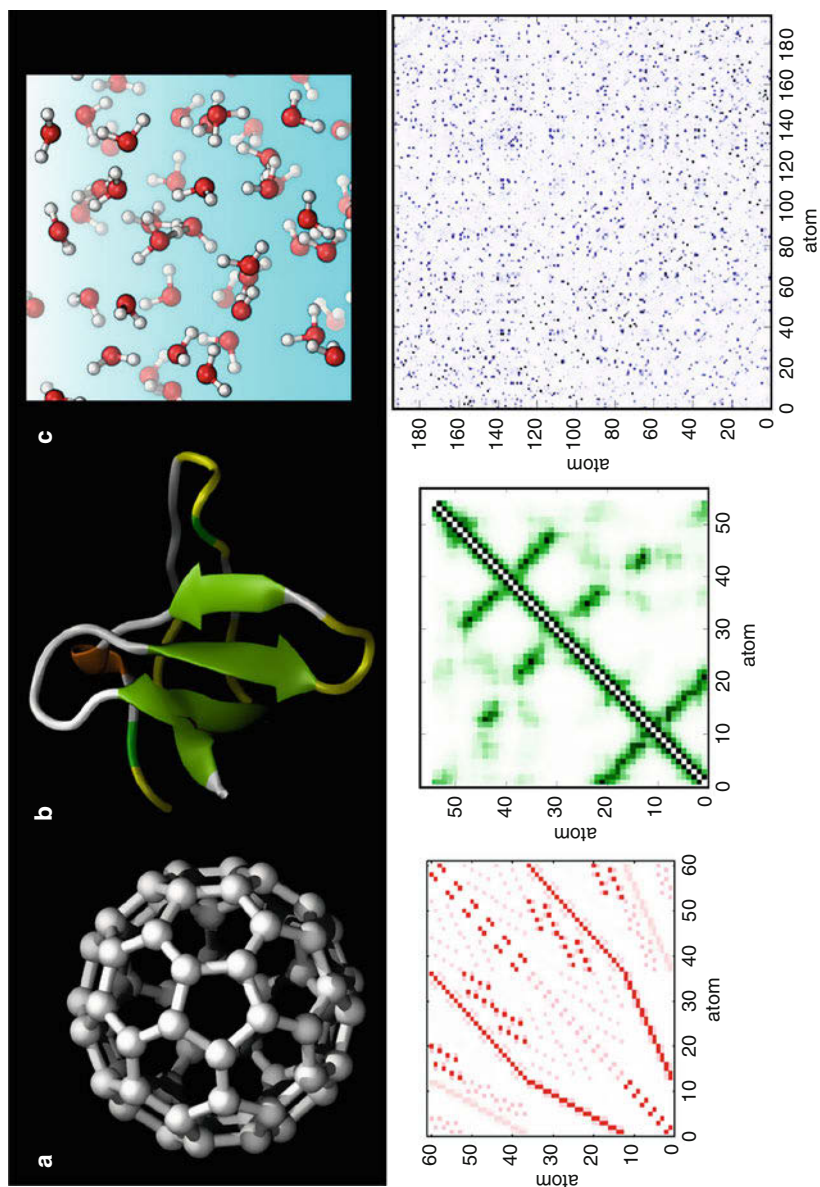
**Fig. 1** Examples of adjacency matrices of (**a**) a fullerene, (**b**) the SH3 protein (using alpha-carbons only), and (**c**) a model of liquid water

continuous changes of atomic structure are reflected in continuous changes of the matrix, and starting from the latter, one can attempt to construct suitable topology-sensitive coordinates, being able to track the transformations of matter in a rather general way. Besides an analysis tool, such coordinates can be also exploited within enhanced sampling methods to accelerate activated processes.

When adopting the adjacency matrix as a basic representation of matter, we are limited by at least two hypotheses. The first is that we can safely neglect large distances between pairs of atoms (where $f$ tends to zero), hoping that they do not play an important role in distinguishing structures or that they are effectively included in the shorter-range network structure. The second hypothesis is that the invariance of $a_{ij}$ under operations leaving the set of all distances unchanged does not introduce ambiguity. Even if counter-examples violating these hypotheses exist – like enantiomer molecules, where a mirror operation may lead to different interactions with biological systems – the vast majority of structural transformations of matter leaves a clear mark in the adjacency matrix. For this reason, several proposals have been made to employ graph theory in order to classify atomic structures and related physicochemical properties (see, e.g., Manolopoulos and Fowler 1992; Ivanciuc and Balaban 1998; Giuliani et al. 2008; García-Domenech et al. 2008; De Corato et al. 2013). In the following sections, graph theory-inspired coordinates are introduced not only for classification purposes but especially to study dynamical transformation pathways.

Note that, in principle, effective CVs correspond to good approximations of the ideal reaction coordinate: the latter can be defined, for the transition between two metastable states $A$ and $B$, by means of the committor function $p_B(\mathbf{R})$ (or equivalently $p_A(\mathbf{R})$) associating each configuration $\mathbf{R}$ in the full $3N$-dimensional space the probability to reach $B$ before $A$ (or equivalently $A$ before $B$). The committor function varies smoothly between zero and one, assuming a value of 0.5 at the transition state; therefore, it indicates the progress of the reaction, and it can be estimated at selected configurations to test the quality of a CV (Bolhuis et al. 2002; Best and Hummer 2005; Weinan and Vanden-Eijnden 2010; Pietrucci 2017).

## 3    Social PeRmutation INvarianT (SPRINT) Coordinates

This section provides a first example of topology-based coordinates: within the sandbox of nanoclusters isomerization, it illustrates the philosophy behind the invention and exploitation of such tools, and it motivates further developments in the direction of phase transitions and chemical reactions. The reader will hopefully be indulgent if here the adjective "social" is attributed, somehow playfully, to coordinates sensitive to the network of neighbors surrounding a given atom.

The physical properties of nanoscale clusters strongly depend on their precise structure, which unfortunately is often out of reach of experimental techniques (Baletto and Ferrando 2005; Billinge and Levin 2007). Simulations can thus play an important role in exploring the thermodynamically relevant configurations. Numerous computational strategies have been applied to determine low-energy

geometries of atomic clusters (Woodley and Catlow 2008; Rossi and Ferrando 2009), including simulated annealing (Ballone et al. 1988; Roethlisberger et al. 1994; Roethlisberger and Andreoni 1991), genetic algorithms (Deaven and Ho 1995; Ho et al. 1998), random search (Lloyd and Johnston 1998; Saunders 2004), basin hopping (Wales and Doye 1997), and minima hopping (Goedecker 2004; Schoenborn and Oganov 2009) methods. As mentioned in the introduction in the case of crystals, usually in all these techniques, the focus is on the fast generation of candidate low-energy structures rather than on the exploration of physically relevant transition pathways.

An alternative approach based on graph theory and on ab initio (density functional theory (DFT)-based) molecular dynamics was introduced in Pietrucci and Andreoni (2011) and exploits the so-called "Social PeRmutation INvarianT" (SPRINT) coordinates. Consider the adjacency matrix $a_{ij}$ of a nanocluster, where the $ij$ are all pairs of atoms: the matrix is symmetric, nonnegative, and also irreducible when it represents a connected graph, i.e., if any pair of vertices is connected through a path. In this case the Perron-Frobenius theorem holds: the largest modulus eigenvalue $\lambda^{\text{max}}$ is real, positive, and nondegenerate, and the corresponding eigenvector $v_i^{\text{max}}$ has all non-zero components with equal sign (we adopt the positive sign convention). In particular, a few very interesting properties can be shown (considering here binary adjacency matrix for simplicity):

- $\lambda^{\text{max}}$ carries global information on the network: it grows with the number of bonds and lies between the average and the maximum coordination number.
- $v_i^{\text{max}}$ carries information about both the short- and long-range topology of the atomic network surrounding atom $i$: for any positive integer $M$

$$v_i^{\text{max}} = \frac{1}{(\lambda^{\text{max}})^M} \sum_j \left(a^M\right)_{ij} v_j^{\text{max}} \tag{1}$$

where $(a^M)_{ij}$ is the number of walks of length $M$ connecting $i$ and $j$. Equation 1 shows the "social character" of $v_i^{\text{max}}$ (Bonacich 1987; Bryan and Leise 2006; Porto et al. 2004).

These observations suggest the possibility to combine the largest eigenvalue and corresponding eigenvector into the following definition of topological SPRINT coordinates:

$$S_i = \sqrt{N} \, \lambda^{\text{max}} \, v_i^{\text{max,sorted}} \, ; \quad i = 1, 2, \ldots, N \tag{2}$$

where $N$ is the number of atoms and the $i$th component must be taken after sorting the eigenvector from its smallest to its largest component. It is this sorting operation that renders the vector **S** invariant with respect to the $N!$ permutations of the labeling of $N$ identical atoms (and thus also with respect to point-group symmetries). Clearly, the constraint $S_1 \leq S_2 \leq \ldots \leq S_N$ strongly reduces the volume of the space to be

**Fig. 2** Examples of simple graphs with each vertex labeled by the SPRINT topological coordinate $S_i$ in Eq. 2. Clearly the coordinate of a vertex is sensitive both to the number of neighbors and to the connectivity of the neighbors themselves
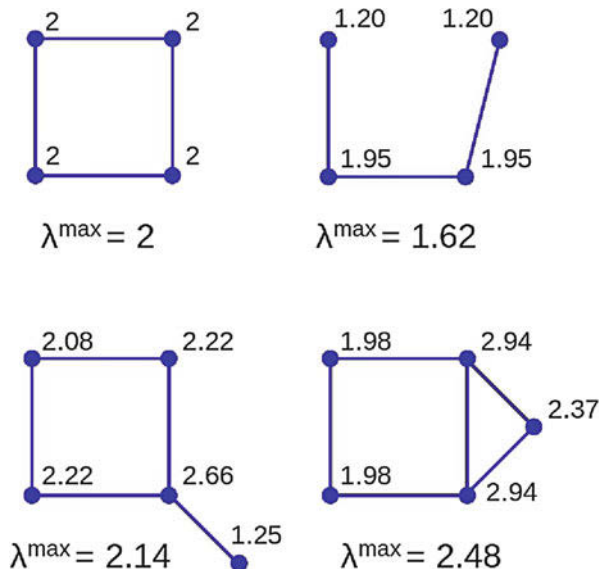
explored: this is a very desirable property in connection with enhanced sampling techniques. On the contrary, a permutation changes the order of rows and columns in $a_{ij}$. An additional advantage is the substantial dimensional reduction from $N(N-1)/2$ elements of the contact matrix to $N$ elements only.

Figure 2 illustrates how the SPRINT coordinates work on a simple example, distinguishing topologically inequivalent atoms and carrying information about both local coordination and the longer-range network topology. In order to track dynamical transitions in a smooth way (as discussed in the previous section), $a_{ij}$ is conveniently generalized from a binary matrix to a real-valued one, by means of a switching function $f(d_{ij})$. Note that the gradual decay of the switching function implies that $S_i$ coordinates contain information not only on the cluster topology but also, to some extent, on the 3D geometry.

It is important to remark that SPRINT coordinates are not a rigorous solution to the problem of resolving isomorphic graphs from truly different ones: in other words, different graphs can share the same SPRINT coordinates. Empirically, however, this degeneracy appears infrequent in small, physically meaningful nanoclusters, where so far these coordinates proved very effective (see below). Instead, different crystal phases sharing a same number of first neighbors for all atoms (e.g., silica) cannot be efficiently resolved by this approach: a solution is presented in next section. It is interesting to note how the principal eigenvector of the adjacency matrix inspired remarkable applications in very different fields, including the compact description of protein structures (Porto et al. 2004), the centrality concept in social networks (Bonacich 1987), and, last but not least, the tremendously successful PageRank algorithm of Google – based on the so-called 25 billion dollars eigenvector (Bryan and Leise 2006).

An interesting application of SPRINT coordinates consists in using them as CVs in combination with metadynamics (Laio and Parrinello 2002; Laio and Gervasio 2008). In the latter technique, a molecular dynamics simulation is supplemented with a history-dependent bias potential constructed as a sum of repulsive Gaussians added in CV space at regular time intervals, with the aim of escaping rapidly from local free-energy minima. SPRINT coordinates are implemented in the plugin PLUMED (Tribello et al. 2014), providing enhanced sampling capabilities to many MD codes. As shown in Pietrucci and Andreoni (2011), in the benchmark case of Lennard-Jones clusters, the simulation, starting from an arbitrary initial structure, is able to quickly explore, among many other isomers, also the global minimum (an icosahedron for 55 atoms and a face-centered truncated octahedron for 38 atoms). An appealing feature of this approach is that the exploration proceeds in a fully blind way, without the need of any educated guess about relevant symmetries or transition mechanisms. Furthermore, the spread of the $S_i$ values easily allows to distinguish high-symmetry from low-symmetry structures: in particular, the passage from a nondegenerate set of SPRINT values to a highly degenerate one marks a disorder-to-order transition. Note that since a large number of CVs are biased at once (one per atom), an unconventional situation in metadynamics applications, the algorithm is somehow in the spirit of local elevation (Huber et al. 1994). The reason is that the aim of this kind of simulations is the quick exploration of a large number of structures with different topologies, not the reconstruction of a low-dimensional free-energy landscape. Alternative CVs for metadynamics simulations of cluster isomerization can be found, e.g., in Tribello et al. (2011) and Rossi and Baletto (2017).

Silicon clusters represent a more realistic system, often taken as benchmark for DFT-based algorithms designed to search for the global minimum. Also for this system, SPRINT-based ab initio metadynamics at room temperature started from arbitrary initial structures lead to the fast and efficient exploration of a multitude of low-energy isomers (Pietrucci and Andreoni 2011). In the case of $Si_{10}$, the consensus lowest-energy structure (tetra-capped trigonal prism) (Ballone et al. 1988; Lyon et al. 2009) is quickly explored, among others. Increasing the cluster size, a richer variety of structures can be observed. For example, in the case of $Si_{16}$, five independent simulations generated hundreds of different geometries over a cumulative time of 400 ps. Geometry optimization identified 16 different isomers within only 45 meV/atom, including a lowest-energy minimum in agreement with previous studies (Goedecker et al. 2005; Yoo and Zeng 2005). Already at this small size, an impressive diversity of structural motifs lie within a narrow energy range, from fused units (e.g., B, C, D, M) to capped (deformed) cores like the tri- or tetra-capped trigonal prism (A, E, F, G, I, K), from symmetric configurations (high SPRINT degeneracy) to quasi-amorphous (low degeneracy), as shown in Fig. 3.

In a study focused on structural transition mechanisms, SPRINT-based ab initio metadynamics was employed to study the spontaneous transformation of small graphene nanoflakes into spheroidal cages (Pietrucci and Andreoni 2014), as observed in transmission electron microscopy experiments (Chuvilin et al. 2010). A set of simulations at different temperatures allowed to identify complex mul-
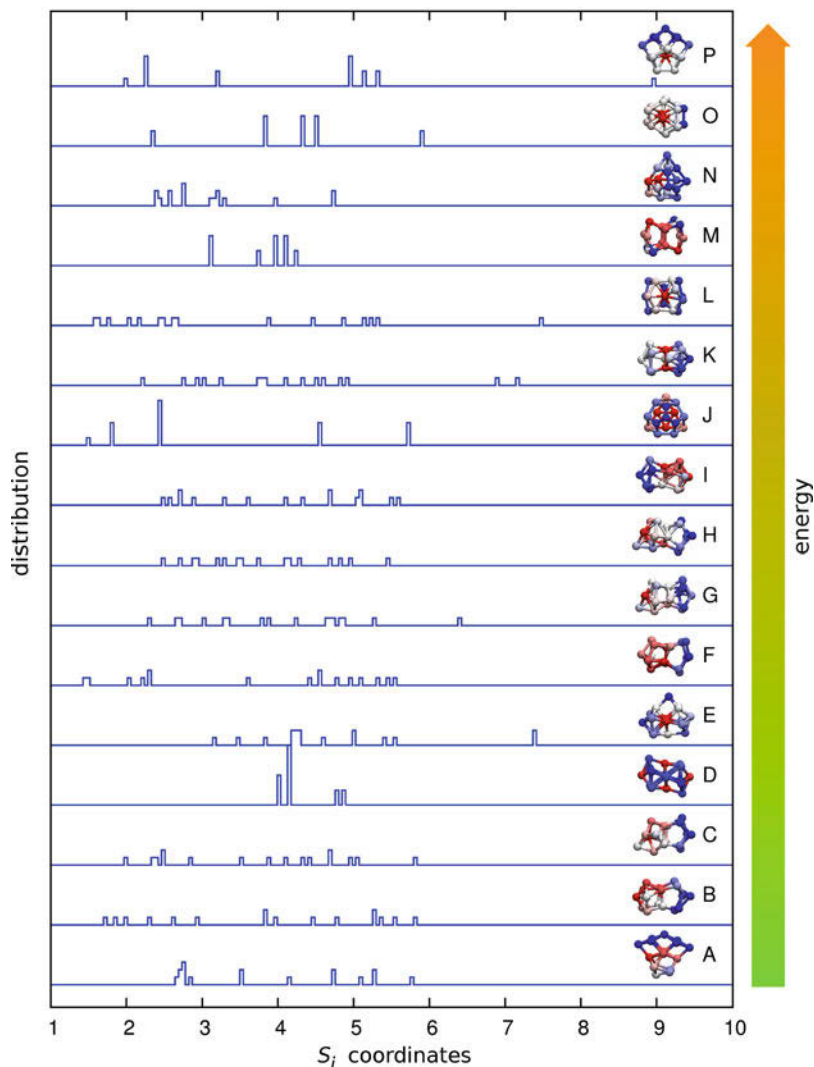
**Fig. 3** SPRINT coordinates distribution for $Si_{16}$ cluster geometries of increasing energy. Atoms are colored from red to blue for increasing magnitude of the corresponding coordinates. The more a structure is symmetric, the higher the level of degeneracy of the coordinates. (Adapted from Pietrucci and Andreoni 2011)

tistep pathways and discovered without any human bias ingenious transformation processes (Fig. 4). Among them, the zipping of a planar flake into a nanocone, the passage from a "bowl" to a cage with the help of carbon chains, and the formation of a pair of pentagons embedded into hexagons (the pyracylene unit, typical of fullerenes) by expelling a dimer from a four-membered ring. The latter
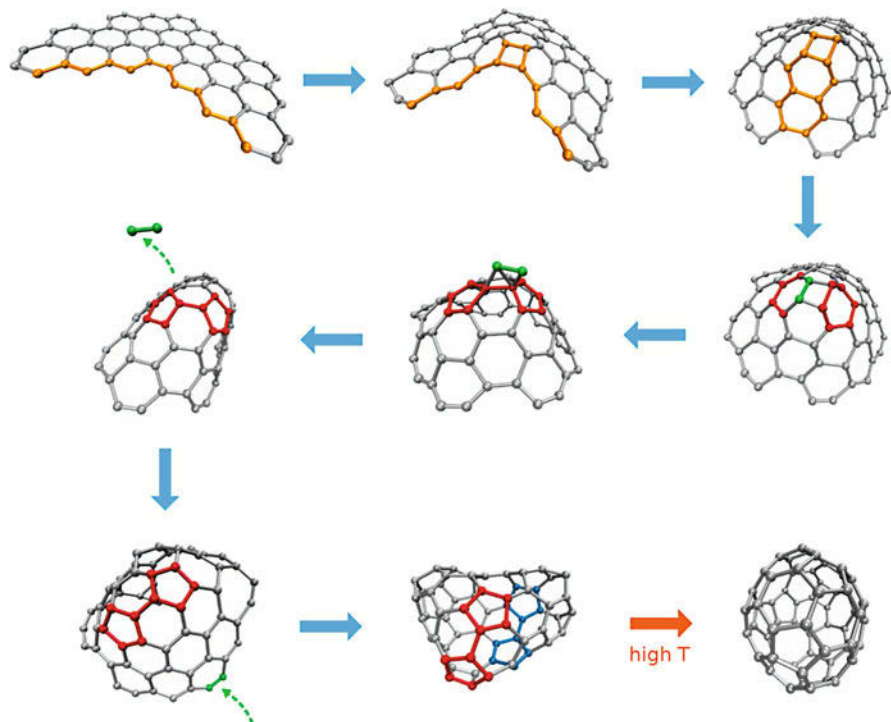
**Fig. 4** Progressive evolution of a graphene flake toward a three-coordinated carbon cage during an ab initio metadynamics simulation based on SPRINT coordinates. Pyracylene units are highlighted in red and blue. (See Pietrucci and Andreoni 2014 for details)

transformation, in particular, is highly nontrivial, pointing to the exploration power of this technique, and is fully compatible with available experiments, being the inverse process of a well-known synthetic route to $C_{62}$ from the $C_{60}$ fullerene (Qian et al. 2003).

A special case of search of cluster structures and related transition mechanisms consists in exploring the possible isomerization reactions of organic molecules: e.g., in Pietrucci and Andreoni (2011), the possible topologies of a $C_4H_5N$ molecule in gas-phase were explored with metadynamics at room temperature. Remarkably, at times the molecule also splits into smaller fragments that are able to recombine if a wall – i.e., a repulsive potential – prevents very large separations, hence probing dissociation and association reactions. Note that if the fragments remain within the range of the switching function, the adjacency matrix does not break down into separate blocks and the principal eigenvector components remain all positive. Note also that, since different elements are present, sorting of the principal eigenvector in Eq. 2 must be performed only within sets of alike atoms. The simulation could explore a sizable list of linear and ring-like isomers, including pyrrole, and smaller molecules like acetylene and hydrogen cyanide. A similar approach, performed

however in a more systematic way (analyzing an ensemble of trajectories), allowed in Zheng and Pfaendtner (2014) to reconstruct a relatively complex reaction network for methanol oxidation, displaying analogies with the consensus network obtained from literature. Other SPRINT applications include the study of $H_2SO_4 \cdot HSO_4^-$ dimer formation in the atmosphere (Wang et al. 2016) and the kinetics of the $S_N2$ reaction of $CH_3Cl + Cl^-$ (Fu et al. 2017). For comparison, the "ab initio nanoreactor" method of Wang et al. (2014) squeezes together molecular species at extreme pressure to discover reaction channels, albeit obtained under unphysical conditions.

An important class of systems is represented by molecules or clusters embedded into a solid matrix: in these cases, in addition to the internal degrees of freedom of the "guest", its possible manifold interactions with the host have also to be included in the configuration space exploration. In Balan et al. (2016), for instance, DFT-based SPRINT-metadynamics simulations allowed to identify the most probable configurations of $B(OH)_4^-$ (an important proxy of past ocean pH) inside calcite and aragonite crystals. Theoretical $^{11}B$ NMR spectra of the predicted structures were found in agreement with experimental results.

Besides enhanced sampling, a different use of topological coordinates aims at the automatic classification of molecular topologies. The basic idea is that different structural isomers are (usually) characterized by different SPRINT vectors that can be directly compared to detect similarity thanks to their invariance under permutation of identical atoms. Examples are the study of complex combustion reactions of hydrocarbons (Lai et al. 2014; Johansson et al. 2016), of the reaction network of $C_3H_4O$ explored with automated transition state search methods (Martínez-Núñez 2015), and of isomers of pure and hydrogenated silicon nanostructures (Baturin et al. 2014) explored with the USPEX structure prediction method (Glass et al. 2006).

On the other hand, as mentioned above, SPRINT coordinates have a limited resolving power in the case of homogeneous bulk systems with uniform coordination patterns, such as pure silica or water phases, for instance. This is not surprising given the dimensional reduction implied in the passage from the adjacency matrix to a $N$-dimensional vector. The next section introduces a different set of topological coordinates that avoid such dimensional reduction, displaying a remarkable ability to distinguish even very similar structures of bulk materials.

## 4 The Difficult Case of Water Poly(a)morphism: A Simple Metric Resolves Different Topologies

Despite the simplicity of its molecular unit, water is a challenging system, because of its uniquely rich polymorphism (Bartels-Rausch et al. 2012), the existence of different amorphous forms (polyamorphism), as well as predicted but yet unconfirmed features, including an elusive liquid-liquid transition (Palmer et al. 2014) and the possible formation of plastic phases (Himoto et al. 2014). It turns out that it is possible to define a simple metric, capturing changes in the topology of the interatomic network as represented by the adjacency matrix $a_{ij}$, being

able to differentiate water (meta)stable structures. The metric further allows to systematically track transitions among liquid, amorphous, and crystalline forms throughout the whole phase diagram of water: this is a crucial feature, because it naturally leads to CVs for enhanced sampling simulations that open the possibility of studying complex transitions like crystal nucleation from liquid or amorphous forms at a cheap computational cost. As shown below, the formulation of the metric and of the related CVs is very general, not specific to water, indicating applicability to a wide range of materials and transformation processes.

Several experiments disclosed connections between stable and metastable water phases (Mishima and Stanley 1998; Klotz et al. 2005), while simulations highlighted the importance of metastable states in understanding the mechanism of phase transitions and related transformations (Russo et al. 2014). A classic example is the connection between the crystalline ice stable at ambient pressure (ice I) and the low-density amorphous (LDA) and high-density amorphous (HDA) ices: by compressing ice I up to 10 kbar at ≈80 K, one obtains HDA instead of ice VI (Mishima et al. 1984), which may be transformed into LDA by decompression of HDA at 130 K (Klotz et al. 2005) or by heating recovered HDA at ambient pressure to beyond 130 K; (Mishima and Stanley 1998) finally ice I is recovered by heating up LDA. Similar connections between crystalline and amorphous ices are found in the high-pressure region of the water phase diagram, where a very-high-density amorphous (VHDA) ice, plastic ices, and crystalline structures with complex hydrogen-bond network (e.g., ice VII) have been observed or predicted (Amann-Winkel et al. 2016; Himoto et al. 2014). Clearly, the experimental evolution of water structures is a complex matter, and the observed outcomes are often determined by kinetics rather than pure thermodynamics. In other words, the evolution strongly depends on the detailed path followed – a trajectory in pressure, temperature, and time variables. Water appears therefore an ideal playground for simulation techniques able to navigate complex free-energy landscapes.

Molecular dynamics, based on realistic interatomic potentials (or, if affordable, DFT) and accelerated by enhanced sampling techniques, is in principle able to track such transitions. However, the CVs suitable to this task are often designed for specific classes of structural transformations (Lechner and Dellago 2008; Martoňák et al. 2003; Giberti et al. 2015; Haji-Akbari and Debenedetti 2015): no general CV scheme was proved successful for a wide class of different problems, in particular those involving disordered systems. Recently, on the other hand, several distance metrics have been developed with the aim of distinguishing and classifying structures of molecular or extended systems, based on atomic environment and/or interatomic networks (Valle and Oganov 2010; Gallet and Pietrucci 2013; Pietrucci and Martoňák 2015; Pietrucci and Saitta 2015; Zhu et al. 2016; De et al. 2016; Piaggi and Parrinello 2017; Martelli et al. 2018; Barthel et al. 2018). In this context, an important question is whether a given metric is able, besides classifying locally stable structures, to also track dynamical transitions in a continuous and accurate way.

A general and effective metric for the comparison of condensed matter structures results from the permutation invariant vector (PIV) introduced in Gallet and
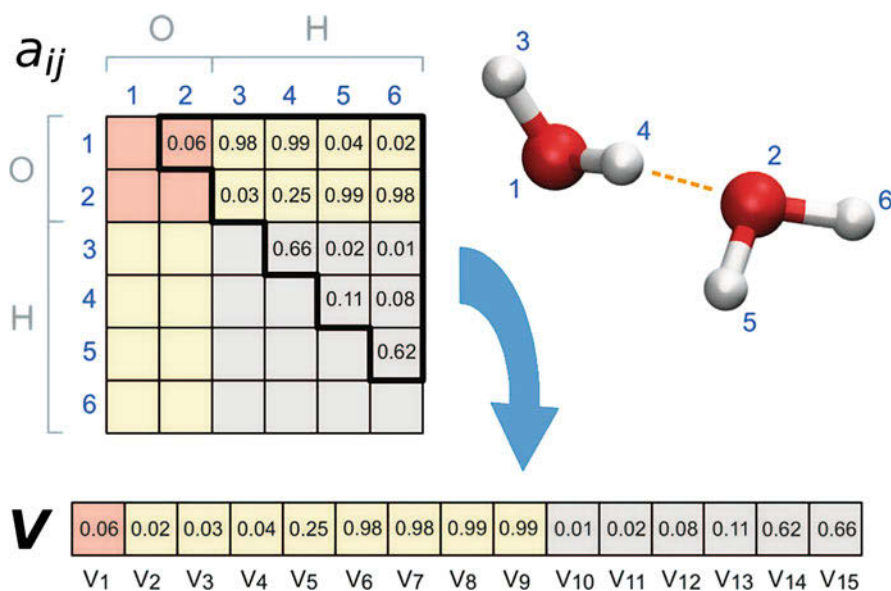
**Fig. 5** Simple system of two water molecules illustrating the definition of PIV: the upper diagonal part of the 6×6 symmetric adjacency matrix $a_{ij}$ (where values close to one correspond to covalent bonds and 0.25 to hydrogen bonds) is rearranged into a vector **V** of 15 components. The entries from each matrix block corresponding to O–O (in red), O–H (in yellow), and H–H pairs (in gray) are sorted in ascending order

Pietrucci (2013) and Pipolo et al. (2017). Each configuration of the system is associated with a vector **V**, while configurations $A$ and $B$ are compared by simply computing the Euclidean distance between the corresponding vectors $||\mathbf{V}^A - \mathbf{V}^B||$. The vector is built by first organizing the elements of a real-valued adjacency matrix into the following blocks $\mathbf{v}_{kk'}$, where $kk'$ indicates a pair of elements (e.g., in the case of water, O–O, O–H or H–H, see Fig. 5):

$$v_{kk'}^{\beta\beta'} = c_{kk'} \ f \left( \sqrt[3]{\frac{\Omega_0}{\Omega}} \ ||\mathbf{R}_{\beta k} - \mathbf{R}_{\beta'k'}|| \right) . \tag{3}$$

where $\mathbf{R}_{\beta k}$ is the position vector of the $\beta$-th atom of type $k$, with $\beta > \beta'$, $k > k'$; $c_{kk'}$ are coefficients that introduce some flexibility (they can be set to one for simplicity); $\Omega$ and $\Omega_0$ are the volume of the simulation box and a reference volume, respectively; and $f$ is, as usual, a switching function monotonically decreasing from one to zero for increasing distance. The PIV **V** is finally obtained by simply sorting the elements of each $\mathbf{v}_{kk'}$ block in ascending ordered and joining together the blocks into a $N(N-1)/2$ -dimensional vector. The sorting operation within each block introduces invariance upon permutation of identical atoms: with growing system size, this operation becomes computationally demanding, so
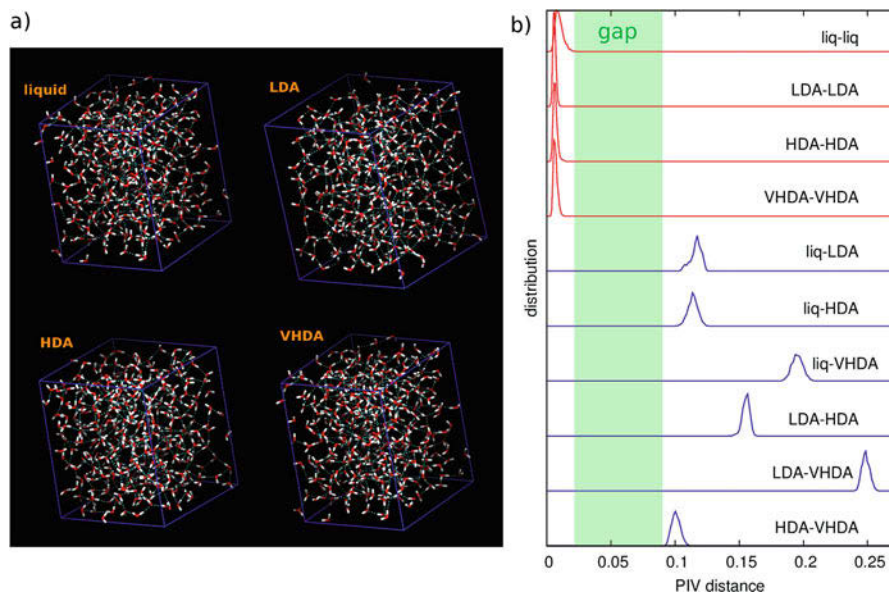
**Fig. 6** (**a**) Representative structures of liquid water and low-density, high-density, and very-high-density amorphous ices. (**b**) Distribution of PIV distances between pairs of structures belonging to different disordered water forms. Only oxygen atoms are included in the definition of **V**, and the switching function $f$ is focused in the range 1.0–4.5 Å. The gap between same-type and different-type distances proves the possibility to automatically resolve the four disordered forms, e.g., by means of cluster analysis. (See Pietrucci and Martoňák 2015 for details)

that efficient parallelization is required for high performance. The volume scaling factor was introduced in Pipolo et al. (2017) to avoid excessive fluctuations of the cell parameters during metadynamics simulations (see below). The choice of $c_{kk'}$ coefficients depends on the application: in water, e.g., the computational burden can often be reduced by focusing on oxygen atoms only with $c_{OO} = 1, c_{OH} = c_{HH} = 0$ (Pietrucci and Martoňák 2015; Pipolo et al. 2017) (see Fig. 6).

The decay range of the switching function is the main parameter entering the PIV definition, which is otherwise very general since it does not depend on specific coordination numbers, angles, symmetries, etc. Conceptually, the switching function $f$ is introduced to focus on a specified range of interatomic distances: short- to medium-range distances are usually the most useful ones, as they display differences between structures (for instance, at the level of pair distribution functions). In water, ranges going from about 2–5 to 2–10 Å allow to resolve different phases (Pipolo et al. 2017): since many neighbors beyond the first shell are included, the term "adjacency matrix" is here employed in a broader sense rather than just considering chemical bonds.

Ideally, the PIV-based metric should display (i) large distances between structures corresponding to different classes like liquid water, crystals, and amorphous forms, (ii) small distances among independent structures within the same class
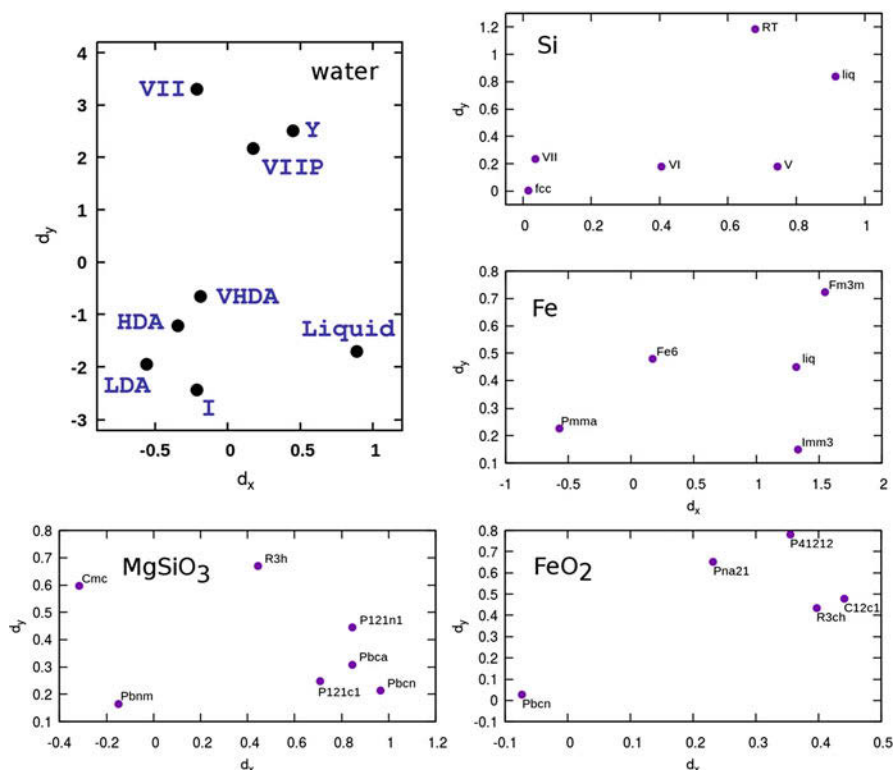
**Fig. 7** Two-dimensional maps faithfully reproducing PIV distances between structures (deviations <6% for water, correlation coefficient >0.99 in the other systems) for a set of diverse materials. The axes are arbitrary. (See Pipolo et al. 2017 for details)

(differing by thermal fluctuations or by the particular realization of the amorphous network). As shown in Fig. 6, this is indeed the case, with a clear gap between distances corresponding to same-type structures and different-type structures. Based on this result, a convenient graphical representation consists in drawing a two-dimensional diagram, or map, where each structure corresponds to a point and where Euclidean distances between PIVs of the different structures are reproduced with good precision (besides this requirement, the axes of the map are arbitrary). As shown in Fig. 7, it is possible (for reasons that remain to be clarified) to draw such maps in a faithful way for different materials, including water, silicon, iron, and oxides, employing a simple Monte Carlo optimization of the positions starting from random initial ones. This type of illustrations is helpful in visualizing the configuration space of different systems (in the same spirit as alternative approaches of Oganov and Valle 2009; De et al. 2016) and suggests that the PIV-based metric is able to systematically distinguish different forms of a material, both ordered and disordered ones, with ample tolerance for the choice of the switching function

range. Further evidence in this direction comes from the analysis of 50 experimental polymorphs belonging to 13 different materials (covalent, metallic, ionic, and molecular), indicating in all cases the effectiveness of the metric (Pipolo et al. 2017).

## 5    Simulating Transitions Between Ordered and Disordered Water Forms

The maps in Fig. 7 are suggestive of interesting possibilities: in particular, one may wonder whether proximity of structures in the map correlates to kinetic proximity, i.e., the existence of a transformation pathway connecting directly the two forms without passing through the others. In the case of water, comparison with the phase diagram and with the known preparation routes of crystalline and amorphous phases suggests a good correlation (Pietrucci and Martoňák 2015). If confirmed for other materials, this correspondence would allow to organize the vast information available about crystal structures (predicted or observed) in terms of possible transition pathways, helping to guide the experimental synthesis of new materials with desirable properties.

From the point of view of enhanced sampling simulations, the PIV-based metric can be employed to construct CVs able to track complex transformation mechanisms, reconstructing the corresponding free-energy landscapes and barriers. A straightforward recipe makes use of the so-called path CVs (Branduardi et al. 2007): given the structure of the system at a given time $\mathbf{R}(t)$ and a set of $n$ reference structures providing a (discretized) putative pathway between an initial state $\mathbf{R}_1$ and a final state $\mathbf{R}_n$, the variables are defined as

$$s(t) = \frac{\sum_{k=1}^{n} k e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)}}{\sum_{k'=1}^{n} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_{k'})}} \tag{4}$$

$$z(t) = -\frac{1}{\lambda} \log \left( \sum_{k=1}^{n} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)} \right) \tag{5}$$

with $s$ quantifying the progress of the transformation and $z$ the distance from the putative path. Both are important: $s$ allows to direct biasing forces to accomplish the transformation, whereas $z$ discriminates between different pathways and allows tracking transitions to states different from the target. Note that the approach shares similarities with the string method (Weinan et al. 2005; Maragliano et al. 2006). The crucial ingredient is the metric $D$ that in the following is taken simply as the squared Euclidean distance between PIVs: $D(\mathbf{R}(t), \mathbf{R}_k) = ||\mathbf{V}(t) - \mathbf{V}_k||^2$ (see Eq. 3). $\lambda$ is a parameter that can be conveniently set to the order of the inverse distance between neighboring reference configurations, assuming they are equally spaced. A much larger $\lambda$ would produce very irregular and discontinuous pathways, while a much smaller one would hamper the resolution of different phases. In several works, path collective variables have been employed to study conformational

changes of biomolecules (Berteotti et al. 2008; Pfaendtner et al. 2009) and chemical reactions in gas phase (Branduardi et al. 2011; Gallet et al. 2012; Pietrucci and Andreoni 2014), based on simple root mean square deviations between Cartesian positions as a metric. Employing PIV-based distances, these coordinates become an effective tool for the simulation of phase transitions. Clearly, if feasible, using only the initial and final states as references ($n = 2$) amounts to a significant simplification, removing the need for educated guesses about the transformation mechanism: application to water, as described below, demonstrates that this can be done, letting the system discover transition pathways without prejudice (Pipolo et al. 2017) (see also Pietrucci and Saitta (2015) for an analogous scheme applied to chemical reactions in solution, using the distance between coordination patterns of selected atoms as a metric).

Figure 8 overlaps the phase diagram of a realistic model of water (TIP4P/2005 Abascal and Vega 2005) with structural transitions obtained by metadynamics and umbrella sampling simulations (at fixed pressure and temperature) based on PIV path CVs. A particularly interesting result is represented by the crystallization transitions both at low and high pressure, since passing from disordered to ordered water forms (particularly without strong supercooling) is probably one of the most difficult tasks in the general field of molecular simulations. At low pressure, the crystallization of ice I is obtained both from the liquid and from amorphous (LDA) ice, in the temperature range 240–260 K around the melting point ($T_m \simeq 250$ K for the adopted interatomic potential Abascal and Vega 2005). As shown in Fig. 9, nucleation of ice proceeds through the formation of a crystal nucleus of cubic symmetry (Ic), ending up in a final state with either a perfect cubic symmetry or



**Fig. 8** The water phase diagram obtained in Abascal and Vega (2005) for the TIP4P/2005 interatomic potential (gray) is superimposed with metadynamics and umbrella sampling simulations of phase transitions (red arrows) between (meta)stable phases (blue). Dashed green lines represent pressure and temperature variations within a phase, obtained with unbiased molecular dynamics simulations. (Adapted from Pipolo et al. 2017)
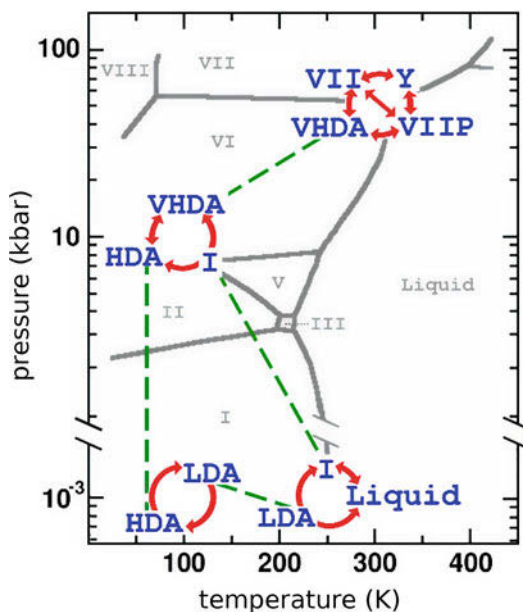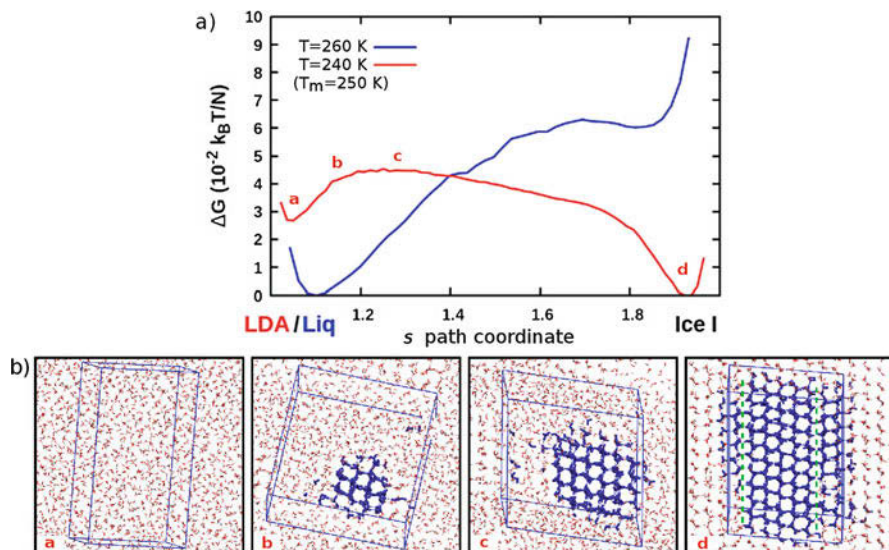
**Fig. 9** (**a**) Free-energy profiles, projected along the $s$ path coordinate, for the crystallization transitions LDA – ice I at $T = 240$ K and liquid – ice I at $T = 260$ K, both at $P = 1$ bar. The latter transition appears thermodynamically unfavorable, consistently with the temperature being higher than the melting temperature of 250 K (TIP4P/2005 potential). Umbrella sampling simulations with $N = 800$ water molecules were seeded from structures extracted from a metadynamics trajectory. The free-energy minima are arbitrarily set to zero. (**b**) Representative structures along the transition pathway for LDA – ice I crystallization. Ice-like molecules, characterized by an averaged local tetrahedral bond order parameter (Lechner and Dellago 2008) of oxygen >0.7, are marked in blue. Snapshot $c$ shows the ice I nucleus at its critical size. Snapshot $d$ displays stacking-disordered ice I, with cubic regions separated by hexagonal layers (green lines). (See Pipolo et al. 2017 for details)

made up of layers of cubic Ic and hexagonal Ih ice, consistently with experiments finding stacking-disordered ice I (Malkin et al. 2012). At high pressure, simulations of the VHDA–ice VII transformation lead to observe two additional metastable configurations: a plastic phase previously proposed in Himoto et al. (2014) and a new phase characterized by a tetragonal oxygen lattice and stacked layers of hydrogen-bond networks. This result illustrates how the approach does not constrain the system to sample configurations along a simple path connecting the two reference structures but rather allows it to follow complex mechanisms and discover new free-energy basins.

The possibility to study crystallization at (or even above) the melting temperature, in the bulk, without any seeds and with the realistic TIP4P/2005 water model indicates that very challenging transformations can be simulated both in favorable and unfavorable thermodynamic conditions, with a moderate computational cost (of the order of 100 ns for metadynamics and of 1000 ns for umbrella sampling simulations) (Pipolo et al. 2017). The PIV-path-CVs approach helped also providing new insight in the problem of heterogeneous ice nucleation, reconstructing

mechanisms, and free-energy landscapes. In Fitzner et al. (2017), ice nucleation was simulated using the mW model of water (Molinero and Moore 2009) and two different fcc model surfaces: the latter give rise to different precritical nucleation clusters that have an important effect on kinetics by controlling which ice polymorph will form. Since different substrates lead to the formation of different ice crystals, heterogeneous classical nucleation theory turns out to be unsuitable to describe the process.

# 6    Conclusions

In the attempt to find a general description of transformation processes at the atomic scale, the adjacency matrix corresponding to the graph of interatomic connections proves a fruitful starting point for structural comparison metrics as well as enhanced sampling coordinates. A distinct feature of this approach is its simple and universal character, as a broad range of condensed matter systems becomes amenable to study in a unified way, addressing ordered and disordered structures on the same footing. In particular, besides the case studies on nanostructures and bulk materials presented in the preceding sections, another important and rapidly growing field of application of such topology-based techniques corresponds to chemical reactions in solution (Pietrucci and Saitta 2015), traditionally difficult to simulate in a robust way as attested by the lack of systematic studies. Among desirable future developments, an important challenge consists in obtaining a more direct access to the kinetic properties of transformation processes (as of today, most enhanced sampling techniques are focused on time-less free-energy landscapes): eventually, the subtle interplay between thermodynamics and kinetics is the key to understand and control the experimental behavior and synthetic routes of materials. In this context, the possibility to construct in a systematic way good approximations of the ideal reaction coordinates represents a significant step forward.

**Index terms** structural transformations, phase transitions, crystallization, nanostructures, disordered systems, chemical reactions, water phase diagram, enhanced sampling, metadynamics, umbrella sampling, free-energy landscapes, structural comparison, structural fingerprints, distance metrics, collective variables, reaction coordinates, graph theory, adjacency matrix, topology of interatomic bond network.

# References

Abascal JL, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 123(23):234505

Amann-Winkel K, Böhmer R, Fujara F, Gainaru C, Geil B, Loerting T (2016) Colloquium: water's controversial glass transitions. Rev Mod Phys 88(1):011002

Balan E, Pietrucci F, Gervais C, Blanchard M, Schott J, Gaillardet J (2016) First-principles study of boron speciation in calcite and aragonite. Geochim Cosmochim Acta 193:119–131

Baletto F, Ferrando R (2005) Structural properties of nanoclusters: energetic, thermodynamic, and kinetic effects. Rev Mod Phys 77:371–423. https://doi.org/10.1103/RevModPhys.77.371

Ballone P, Andreoni W, Car R, Parrinello M (1988) Equilibrium structures and finite temperature properties of silicon microclusters from ab initio molecular-dynamics calculations. Phys Rev Lett 60:271–274

Bartels-Rausch T, Bergeron V, Cartwright JH, Escribano R, Finney JL, Grothe H, Gutiérrez PJ, Haapala J, Kuhs WF, Pettersson JB et al (2012) Ice structures, patterns, and processes: a view across the icefields. Rev Mod Phys 84(2):885

Barthel S, Alexandrov EV, Proserpio DM, Smit B (2018) Distinguishing metal-organic frameworks. Cryst Growth Des 18(3):1738–1747. https://doi.org/10.1021/acs.cgd.7b01663

Baturin V, Lepeshkin S, Magnitskaya M, Matsko N, Uspenskii YA (2014) Structural and electronic properties of small silicon clusters. J Phys Conf Ser 510:012032

Berteotti A, Cavalli A, Branduardi D, Gervasio FL, Recanatini M, Parrinello M (2008) Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations. J Am Chem Soc 131(1):244–250

Best RB, Hummer G (2005) Reaction coordinates and rates from transition paths. Proc Natl Acad Sci USA 102(19):6732–6737

Billinge SJL, Levin I (2007) The problem with determining atomic structure at the nanoscale. Science 316(5824):561–565

Bolhuis P, Chandler D, Dellago C, Geissler P (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annu Rev Phys Chem 53:291–318

Bonacich P (1987) Power and centrality: a family of measures. Am J Sociol 92(5):1170–1182

Branduardi D, Gervasio FL, Parrinello M (2007) From a to b in free energy space. J Chem Phys 126(5):054,103

Branduardi D, De Vivo M, Rega N, Barone V, Cavalli A (2011) Methyl phosphate dianion hydrolysis in solution characterized by path collective variables coupled with DFT-based enhanced sampling simulations. J Chem Theory Comput 7(3):539–543

Bryan K, Leise T (2006) The \$25,000,000,000 eigenvector: the linear algebra behind google. SIAM Rev 48(3):569–581

Chuvilin A, Kaiser U, Bichoutskaia E, Besley NA, Khlobystov AN (2010) Direct transformation of graphene to fullerene. Nat Chem 2(6):450–453

De S, Bartók AP, Csányi G, Ceriotti M (2016) Comparing molecules and solids across structural and alchemical space. Phys Chem Chem Phys 18(20):13754–13769

Deaven D, Ho K (1995) Molecular geometry optimization with a genetic algorithm. Phys Rev Lett 75:288

De Corato M, Bernasconi M, D'Alessio L, Ori O, Putz MV, Benedek G (2013) Topological versus physical and chemical properties of negatively curved carbon surfaces. In: Ashrafi AR, Cataldo F, Iranmanesh A, Ori O (eds) Topological modelling of nanostructures and extended systems. Springer, Dordrecht, pp 105–136

Fitzner M, Sosso GC, Pietrucci F, Pipolo S, Michaelides A (2017) Pre-critical fluctuations and what they disclose about heterogeneous crystal nucleation. Nat Commun 8(1):2257

Fu CD, Oliveira LF, Pfaendtner J (2017) Assessing generic collective variables for determining reaction rates in metadynamics simulations. J Chem Theory Comput 13(3):968–973

Gallet GA, Pietrucci F (2013) Structural cluster analysis of chemical reactions in solution. J Chem Phys 139(7):074,101

Gallet G, Pietrucci F, Andreoni W (2012) Bridging static and dynamical descriptions of chemical reactions: an ab initio study of $CO_2$ interacting with water molecules. J Chem Theory Comput 8(11):4029–4039. https://doi.org/10.1021/ct300581n

García-Domenech R, Gálvez J, de Julián-Ortiz JV, Pogliani L (2008) Some new trends in chemical graph theory. Chem Rev 108(3):1127–1169

Giberti F, Salvalaglio M, Parrinello M (2015) Metadynamics studies of crystal nucleation. IUCrJ 2(2):256–266

Giuliani A, Krishnan A, Zbilut JP, Tomita M (2008) Proteins as networks: usefulness of graph theory in protein science. Curr Protein Pept Sci 9(1):28–38

Glass CW, Oganov AR, Hansen N (2006) Uspex an evolutionary crystal structure prediction. Comput Phys Commun 175(11):713–720

Goedecker S (2004) Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. J Chem Phys 120:9911–9917

Goedecker S, Hellmann W, Lenosky T (2005) Global minimum determination of the Born-Oppenheimer surface within density functional theory. Phys Rev Lett 95(5):055501

Haji-Akbari A, Debenedetti PG (2015) Direct calculation of ice homogeneous nucleation rate for a molecular model of water. Proc Natl Acad Sci USA 112(34):10582–10588

Himoto K, Matsumoto M, Tanaka H (2014) Yet another criticality of water. Phys Chem Chem Phys 16(11):5081–5087

Ho K, Shvartsburg A, Pan B, Lu Z, Wang C, Wacker J, Fye J, Jarrold M (1998) Structures of medium-sized silicon clusters. Nature 392:582

Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aid Mol Des 8(6):695–708

Ivanciuc O, Balaban AT (1998) Graph theory in chemistry. In: Schleyer PVR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III HF, Schreiner PR (eds) The encyclopedia of computational chemistry. John Wiley & Sons, Chichester, pp 1169–1190

Johansson KO, Dillstrom T, Monti M, El Gabaly F, Campbell MF, Schrader PE, Popolan-Vaida DM, Richards-Henderson NK, Wilson KR, Violi A et al (2016) Formation and emission of large furans and oxygenated hydrocarbons from flames. Proc Natl Acad Sci USA 113:8374–8379

Klotz S, Strassle T, Nelmes R, Loveday J, Hamel G, Rousse G, Canny B, Chervin J, Saitta A (2005) Nature of the polyamorphic transition in ice under pressure. Phys Rev Lett 94:025506

Lai JY, Elvati P, Violi A (2014) Stochastic atomistic simulation of polycyclic aromatic hydrocarbon growth in combustion. Phys Chem Chem Phys 16(17):7969–7979

Laio A, Gervasio FL (2008) Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys 71(12):126601

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99(20):12562–12566

Lechner W, Dellago C (2008) Accurate determination of crystal structures based on averaged local bond order parameters. J Chem Phys 129(11):114707

Lloyd L, Johnston RL (1998) Modelling aluminium clusters with an empirical many-body potential. Chem Phys 236:107

Lyon JT, Gruene P, Fielicke A, Meijer G, Janssens E, Claes P, Lievens P (2009) Structures of silicon cluster cations in the gas phase. J Am Chem Soc 131(3):1115–1121

Malkin TL, Murray BJ, Brukhno AV, Anwar J, Salzmann CG (2012) Structure of ice crystallized from supercooled water. Proc Natl Acad Sci USA 109(4):1041–1045

Manolopoulos DE, Fowler PW (1992) Molecular graphs, point groups, and fullerenes. J Chem Phys 96(10):7603–7614

Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) String method in collective variables: minimum free energy paths and isocommittor surfaces. J Chem Phys 125(2):024106

Martelli F, Ko HY, Oğuz EC, Car R (2018) Local-order metric for condensed-phase environments. Phys Rev B 97(6):064105

Martínez-Núñez E (2015) An automated transition state search using classical trajectories initialized at multiple minima. Phys Chem Chem Phys 17(22):14912–14921

Martoňák R, Laio A, Parrinello M (2003) Predicting crystal structures: the Parrinello-Rahman method revisited. Phys Rev Lett 90(7):075503

Mishima O, Stanley HE (1998) The relationship between liquid, supercooled and glassy water. Nature 396(6709):329–335

Mishima O, Calvert L, Whalley E (1984) Melting ice I at 77 K and 10 kbar: a new method of making amorphous solids. Nature 310(5976):393–395

Molinero V, Moore EB (2009) Water modeled as an intermediate element between carbon and silicon. J Phys Chem B 113(13):4008–4016. https://doi.org/10.1021/jp805227c

Oganov AR, Valle M (2009) How to quantify energy landscapes of solids. J Chem Phys 130(10):104504

Palmer JC, Martelli F, Liu Y, Car R, Panagiotopoulos AZ, Debenedetti PG (2014) Metastable liquid-liquid transition in a molecular model of water. Nature 510(7505):385–388

Pfaendtner J, Branduardi D, Parrinello M, Pollard TD, Voth GA (2009) Nucleotide-dependent conformational states of actin. Proc Natl Acad Sci USA 106(31):12723–12728

Piaggi PM, Parrinello M (2017) Entropy based fingerprint for local crystalline order. J Chem Phys 147(11):114112

Pickard CJ, Needs R (2011) Ab initio random structure searching. J Phys Condens Matt 23(5):053201

Pietrucci F (2017) Strategies for the exploration of free energy landscapes: unity in diversity and challenges ahead. Rev Phys 2:32–45

Pietrucci F, Andreoni W (2011) Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale. Phys Rev Lett 107:085504. https://doi.org/10.1103/PhysRevLett.107.085504

Pietrucci F, Andreoni W (2014) Fate of a graphene flake: a new route toward fullerenes disclosed with ab initio simulations. J Chem Theory Comput 10(3):913–917

Pietrucci F, Martoňák R (2015) Systematic comparison of crystalline and amorphous phases: charting the landscape of water structures and transformations. J Chem Phys 142(10):104704

Pietrucci F, Saitta AM (2015) Formamide reaction network in gas phase and solution via a unified theoretical approach: toward a reconciliation of different prebiotic scenarios. Proc Natl Acad Sci USA 112(49):15030–15035

Pipolo S, Salanne M, Ferlat G, Klotz S, Saitta AM, Pietrucci F (2017) Navigating at will on the water phase diagram. Phys Rev Lett 119(24):245701

Porto M, Bastolla U, Roman H, Vendruscolo M (2004) Reconstruction of protein structures from a vectorial representation. Phys Rev Lett 92(21):218101

Qian W, Chuang SC, Amador RB, Jarrosson T, Sander M, Pieniazek S, Khan SI, Rubin Y (2003) Synthesis of stable derivatives of $C_{62}$: the first nonclassical fullerene incorporating a four-membered ring. J Am Chem Soc 125(8):2066–2067. https://doi.org/10.1021/ja029679s

Radha A, Lander L, Rousse G, Tarascon J, Navrotsky A (2015) Thermodynamic stability and correlation with synthesis conditions, structure and phase transformations in orthorhombic and monoclinic Li2M(SO4)2 (M = Mn,Fe,Co,Ni) polymorphs. J Mater Chem A 3(6):2601–2608

Roethlisberger U, Andreoni W (1991) Structural and electronic-properties of sodium microclusters (n = 2−20) at low and high temperatures: New insights from ab initio molecular-dynamics studies. J Chem Phys 94(12):8129–8151

Roethlisberger U, Andreoni W, Parrinello M (1994) Structure of nanoscale silicon clusters. Phys Rev Lett 72:665–668

Rossi K, Baletto F (2017) The effect of chemical ordering and lattice mismatch on structural transitions in phase segregating nanoalloys. Phys Chem Chem Phys 19(18):11057–11063

Rossi G, Ferrando R (2009) Searching for low-energy structures of nanoparticles: a comparison of different methods and algorithms. J Phys Condens Matter 21(8):084208

Russo J, Romano F, Tanaka H (2014) New metastable form of ice and its role in the homogeneous crystallization of water. Nat Mater 13(7):733–739

Saunders M (2004) Stochastic search for isomers on a quantum mechanical surface. J Comput Chem 25(5):621–626

Schoenborn SE, Goedecker S, Roy S, Oganov AR (2009) The performance of minima hopping and evolutionary algorithms for cluster structure prediction. J Chem Phys 130:144108

Schreiber RE, Houben L, Wolf SG, Leitus G, Lang ZL, Carbó JJ, Poblet JM, Neumann R (2017) Real-time molecular scale observation of crystal formation. Nat Chem 9:369–373

Tribello GA, Cuny J, Eshet H, Parrinello M (2011) Exploring the free energy surfaces of clusters using reconnaissance metadynamics. J Chem Phys 135(11):114109

Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) Plumed 2: new feathers for an old bird. Comput Phys Commun 185(2):604–613

Valle M, Oganov AR (2010) Crystal fingerprint space–a novel paradigm for studying crystal-structure sets. Acta Cryst Sect A 66(5):507–517

Wales D, Doye J (1997) Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 101:5111–5116

Wang LP, Titov A, McGibbon R, Liu F, Pande VS, Martínez TJ (2014) Discovering chemistry with an ab initio nanoreactor. Nat Chem 6(12):1044–1048

Wang Y, Lv J, Zhu L, Ma Y (2012) Calypso: a method for crystal structure prediction. Comput Phys Commun 183(10):2063–2070

Wang Y, Huang Y, Gu B, Xiao X, Liang D, Rao W (2016) Formation of the $H_2SO_4 \cdot HSO_4^-$ dimer in the atmosphere as a function of conditions: a simulation study. Mol Phys 114(23):3475–3482

Weinan E, Vanden-Eijnden E (2010) Transition-path theory and path-finding algorithms for the study of rare events. Ann Rev Phys Chem 61:391–420

Weinan E, Ren W, Vanden-Eijnden E (2005) Finite temperature string method for the study of rare events. J Phys Chem B 109(14):6688–6693. https://doi.org/10.1021/jp0455430

Wilmer CE, Leaf M, Lee CY, Farha OK, Hauser BG, Hupp JT, Snurr RQ (2012) Large-scale screening of hypothetical metal–organic frameworks. Nat Chem 4(2):83–89

Woodley S, Catlow R (2008) Crystal structure prediction from first principles. Nat Mater 7:937–946

Yoo S, Zeng XC (2005) Structures and stability of medium-sized silicon clusters. III. Reexamination of motif transition in growth pattern from Si15 to Si20. J Chem Phys 123(16):164303

Zheng S, Pfaendtner J (2014) Car–Parrinello molecular dynamics+ metadynamics study of high-temperature methanol oxidation reactions using generic collective variables. J Phys Chem C 118(20):10764–10770

Zhu L, Amsler M, Fuhrer T, Schaefer B, Faraji S, Rostami S, Ghasemi SA, Sadeghi A, Grauzinyte M, Wolverton C, Goedecker S (2016) A fingerprint based metric for measuring similarities of crystalline structures. J Chem Phys 144(3):034203

# Variationally Enhanced Sampling

# 29

Omar Valsson  and Michele Parrinello

## Contents

O. Valsson
Max Planck Institute for Polymer Research, Mainz, Germany
e-mail: valsson@mpip-mainz.mpg.de

M. Parrinello (✉)
Department of Chemistry and Applied Biosciences, ETH Zurich, Lugano, Switzerland

Facoltà di Informatica, Instituto di Scienze Computationali, Università della Svizzera italiana,
Lugano, Switzerland

Istituto Italiano di Tecnologia, Genova, Italy
e-mail: parrinello@phys.chem.ethz.ch

**Abstract**

Atomistic simulations can give a microscopic understanding of materials. However, their use is generally limited by the time scales that can be accessed as normally this falls short of what is needed to properly sample the complex free energy landscapes that characterize most material systems. Advanced methods are thus needed to enhance the sampling and overcome this time scale problem. Variationally enhanced sampling is one such approach based on identifying the important slow degrees of freedom and enhancing their fluctuations through the introduction of an external bias potential. The method is based on a variational principle that shows how the bias potential can be constructed by minimizing a convex functional, which brings a lot of flexibility. We introduce the theory and methodology of variationally enhanced sampling and discuss various novel and innovative applications and extensions.

## 1    Introduction

Atomistic simulations constitute an important part of material modeling allowing for a microscopic understanding of their behavior. However, it is well known that such simulations generally suffer from severe limitations as the time scales that can be accessed fall short of what is needed. A case in point is nucleation of a crystal, in most cases the nucleation process occurs on much longer time scales than can be simulated in practice. This comes from the fact that free energy landscapes of material systems are characterized by many metastable states that are separated by barriers much larger than the thermal energy, leading to kinetic bottlenecks. This problem is of course quite general and comes up anywhere atomistic simulations are used in chemistry, physics, and biology. It is therefore of great importance to develop advanced methods that lead to an enhanced sampling of phase space.

A large and important class of such enhanced sampling techniques is based on identifying the important slow degrees of freedom and in different manners enhancing their fluctuations, allowing the system more easily to migrate between metastable states (Valsson et al. 2016). In practice this is done by considering so-called collective variables (CVs) which are coarse-grained descriptors that depend on the atomic coordinates and properly distinguish between the metastable states of the system. CV fluctuations are then enhanced by adding a bias potential that acts in the space spanned by the CVs. The origin of this idea can be traced to umbrella sampling introduced by Torrie and Valleau (1977) where they employed a static bias potential obtained by trial and error. Constructing a suitable bias potential that effectively enhances CV fluctuations is far from straightforward as this implies knowledge about the underlying free energy landscape which is nearly always unknown a priori and the very object that one is interested in obtaining. Methods have thus been developed that construct the bias on-the-fly during the simulation. Among such techniques are, for example, local elevation (Huber et al. 1994), adaptive biasing force (Darve and Pohorille 2001), Wang-Landau algorithm

(Wang and Landau 2001), energy landscape paving (Hansmann and Wille 2002), Gaussian-mixture umbrella sampling (Maragakis et al. 2009), and metadynamics (Laio and Parrinello 2002; Barducci et al. 2008), to name a few.

Variationally enhanced sampling (VES) (Valsson and Parrinello 2014) is a method in this spirit that we introduced in 2014. What makes VES interesting is that it is based on a variational principle that shows how an effective bias potential can be constructed by minimizing a convex functional. This brings a lot of flexibility, for example, in the form of the bias potential that can be employed. Furthermore, it is possible to tailor in which way the CV fluctuations are enhanced. We will here introduce the theory and methodology of VES and discuss its various applications and extensions.

## 2 Theory and Methodology of Variationally Enhanced Sampling

### 2.1 Collective Variable-Based Enhanced Sampling

We consider a system described by coordinates $\mathbf{R}$ and potential energy function $U(\mathbf{R})$. The Boltzmann distribution, which one wants to sample with molecular dynamics (MD) or Monte Carlo simulations at a given temperature $T$, is defined as $P(\mathbf{R}) = e^{-\beta U(\mathbf{R})}/\int d\mathbf{R}\, e^{-\beta U(\mathbf{R})}$ where $\beta = (k_\mathrm{B}T)^{-1}$ is the inverse temperature. One can introduce some set of collective variables $\mathbf{s}(\mathbf{R}) = (s_1(\mathbf{R}), s_2(\mathbf{R}), \ldots, s_n(\mathbf{R}))$ that depend on $\mathbf{R}$ and are able to properly distinguish between the relevant metastable states. The equilibrium probability distribution associated with the CVs is defined as

$$P(\mathbf{s}) = \langle \delta\left(\mathbf{s} - \mathbf{s}(\mathbf{R})\right)\rangle = \int d\mathbf{R}\, \delta(\mathbf{s} - \mathbf{s}(\mathbf{R}))\, P(\mathbf{R}). \tag{1}$$

The free energy surface (FES) associated with the CVs is defined as the logarithm of the probability distribution,

$$F(\mathbf{s}) = -\frac{1}{\beta} \log P(\mathbf{s}), \tag{2}$$

where we can ignore immaterial constants. Under the influence of an external bias potential $V(\mathbf{s}(\mathbf{R}))$ that acts in the space spanned by the CVs, the biased probability distribution is

$$P_V(\mathbf{s}) = \langle \delta\left(\mathbf{s} - \mathbf{s}(\mathbf{R})\right)\rangle_V = \int d\mathbf{R}\, \delta(\mathbf{s} - \mathbf{s}(\mathbf{R}))\, P_V(\mathbf{R})$$

$$= \frac{e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}} \propto P(\mathbf{s})\, e^{-\beta V(\mathbf{s})} \tag{3}$$

where $P_V(\mathbf{R}) = e^{-\beta[U(\mathbf{R})+V(\mathbf{s}(\mathbf{R}))]} / \int d\mathbf{R}\, e^{-\beta[U(\mathbf{R})+V(\mathbf{s}(\mathbf{R}))]}$ is the biased Boltzmann distribution.

## 2.2 Variational Principle to Enhanced Sampling

Variationally enhanced sampling is based on a variational principle given by a functional $\Omega[V]$ that depends on an external bias potential $V(\mathbf{s})$ (Valsson and Parrinello 2014). The functional is defined as

$$\Omega[V] = \frac{1}{\beta} \log \frac{\int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s}\, e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s}\, p(\mathbf{s}) V(\mathbf{s}), \tag{4}$$

where $p(\mathbf{s})$ is a predefined probability distribution that is assumed to be normalized. It can be shown that $\Omega[V]$ is invariant under the addition of a constant to $V(\mathbf{s})$, namely, $\Omega[V + k] = \Omega[V]$, and that it is a convex functional, i.e., fulfills

$$\Omega\left[\frac{V_1 + V_2}{2}\right] \leq \frac{1}{2}\Omega[V_1] + \frac{1}{2}\Omega[V_2]. \tag{5}$$

The minimum of $\Omega[V]$ is given up to a constant by

$$V(\mathbf{s}) = -F(\mathbf{s}) - \frac{1}{\beta} \log p(\mathbf{s}). \tag{6}$$

It follows from the convexity of $\Omega[V]$ that this is the global minimum. By inserting this bias potential into Eq. (3), which gives $P_V(\mathbf{s}) = p(\mathbf{s})$, one can see that at the minimum the CVs are sampled according to the distribution $p(\mathbf{s})$. In other words, the distribution $p(\mathbf{s})$ is a so-called target distribution that determines the sampling of CVs that is obtained when minimizing $\Omega[V]$.

The variational principle implies that one can construct an effective bias potential by minimizing $\Omega[V]$ and that the FES $F(\mathbf{s})$ can be directly obtained from the bias potential that minimizes $\Omega[V]$. Furthermore, by appropriately choosing the target distribution, one can tailor the way in which CV fluctuations are enhanced. As discussed below there is considerable freedom in how this can be done.

The functional $\Omega[V]$ has close relation to the concepts of cross entropy, relative entropy, and the Kullback-Leibler (KL) divergence (Rubinstein 1999; Shell 2008; Bilionis and Koutsourelakis 2012; Zhang et al. 2014). This can be seen by rewriting the functional as (Invernizzi et al. 2017)

$$\beta\Omega[V] = D_{\mathrm{KL}}(p||P_V) - D_{\mathrm{KL}}(p||P) = H_{\mathrm{x}}(p||P_V) - H_{\mathrm{x}}(p||P), \tag{7}$$

where $D_{\mathrm{KL}}(P||Q) = \int d\mathbf{s}\, P(\mathbf{s}) \log \frac{P(\mathbf{s})}{Q(\mathbf{s})}$ is the KL divergence (or relative entropy) between two probability distributions $P(\mathbf{s})$ and $Q(\mathbf{s})$ and $H_{\mathrm{x}}(P||Q) =$

$- \int d\mathbf{s}\, P(\mathbf{s}) \log Q(\mathbf{s})$ is the cross entropy. Thus, minimizing $\Omega[V]$ is equivalent to minimizing the KL divergence or the cross entropy between the biased distribution $P_V(\mathbf{s})$ and the target distribution $p(\mathbf{s})$.

## 2.3 Minimization in Practice

In practical applications the variational principle is utilized by assuming some functional form for $V(\mathbf{s}; \boldsymbol{\alpha})$ that depends on a set of variational parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$. In this way, one goes from an abstract functional minimization to that of the multidimensional function $\Omega(\boldsymbol{\alpha}) = \Omega[V(\boldsymbol{\alpha})]$. During the minimization of $\Omega(\boldsymbol{\alpha})$, the variational parameters $\boldsymbol{\alpha}$ are iteratively updated by using some recursion formula that utilizes the gradient $\nabla \Omega(\boldsymbol{\alpha})$. The elements of the gradient $\nabla \Omega(\boldsymbol{\alpha})$ are defined as

$$\frac{\partial \Omega(\boldsymbol{\alpha})}{\partial \alpha_i} = - \left\langle \frac{\partial V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i} \right\rangle_{V(\boldsymbol{\alpha})} + \left\langle \frac{\partial V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i} \right\rangle_p, \tag{8}$$

where the expectation values are obtained under the influence of the bias potential $V(\mathbf{s}; \boldsymbol{\alpha})$ or over the target distribution $p(\mathbf{s})$, respectively. If needed, the Hessian matrix $\mathbf{H}(\boldsymbol{\alpha})$ can be defined in similar manner as

$$\frac{\partial^2 \Omega(\boldsymbol{\alpha})}{\partial \alpha_i\, \partial \alpha_j} = - \left\langle \frac{\partial^2 V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i\, \partial \alpha_j} \right\rangle_{V(\boldsymbol{\alpha})} + \left\langle \frac{\partial^2 V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i\, \partial \alpha_j} \right\rangle_p$$

$$+ \beta \operatorname{cov} \left[ \frac{\partial V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_i}, \frac{\partial V(\mathbf{s}; \boldsymbol{\alpha})}{\partial \alpha_j} \right]_{V(\boldsymbol{\alpha})}. \tag{9}$$

Since $\nabla \Omega(\boldsymbol{\alpha})$ is intrinsically noisy due to statistical averaging, it is necessary to use stochastic optimization methods. In particular the averaged stochastic gradient descent algorithm introduced by Bach and Moulines (2013) has proven a good choice so far. In this algorithm the variational parameters $\boldsymbol{\alpha}$ are updated according to the following recursion equation

$$\boldsymbol{\alpha}^{(n+1)} = \boldsymbol{\alpha}^{(n)} - \mu \left[ \nabla \Omega(\bar{\boldsymbol{\alpha}}^{(n)}) + \mathbf{H}(\bar{\boldsymbol{\alpha}}^{(n)}) \left[ \boldsymbol{\alpha}^{(n)} - \bar{\boldsymbol{\alpha}}^{(n)} \right] \right], \tag{10}$$

where $\mu$ is a fixed step size and the gradient and Hessian $\mathbf{H}(\bar{\boldsymbol{\alpha}}^{(n)})$ are obtained using the averaged parameters $\bar{\boldsymbol{\alpha}}^{(n)} = \frac{1}{n+1} \sum_{k=0}^{n} \boldsymbol{\alpha}^{(k)}$ (i.e., the bias potential depends on the averaged parameters). The averaging leads to a smooth convergence of the bias and the estimated FES. Furthermore, it allows for the usage of rather short sampling time for each iteration, on the order of few thousand MD steps. It should be noted that this averaged stochastic gradient descent algorithm is not a unique choice and other stochastic optimization methods could be used. In fact, it is an interesting

research question to develop and benchmark new optimization algorithm to perform the minimization of $\Omega(\boldsymbol{\alpha})$.

It has proven quite useful to utilize multiple walkers (Raiteri et al. 2006) within VES. In this scheme multiple copies of the system that share the same bias potential are run in parallel. The averages needed for the gradient and the Hessian are then obtained by combined sampling of the different copies. Furthermore, in the same way as for metadynamics (Bussi et al. 2006), VES can be used together with parallel tempering or other replica exchange techniques that help with sampling missing slow degrees of freedom.

## 2.4    Linear Basis Set Expansion

The most natural way to represent the bias potential is to consider a linear expansion in some set of basis functions $f_k(\mathbf{s})$,

$$V(\mathbf{s}; \boldsymbol{\alpha}) = \sum_k \alpha_k \, f_k(\mathbf{s}). \tag{11}$$

Most applications so far have employed basis functions constructed as tensor product of one-dimensional orthogonal basis functions such as Fourier series (i.e., plane waves) for periodic CVs and Legendre or Chebyshev polynomials for nonperiodic CVs. For example, for two CVs the bias potential is written as

$$V(s_1, s_2; \boldsymbol{\alpha}) = \sum_{i,j} \alpha_{i,j} \, g_i(s_1) \, h_j(s_2), \tag{12}$$

where $g_i(s_1)$ and $h_j(s_2)$ are some one-dimensional orthogonal basis functions. A relatively modest number of basis functions is needed in most cases as FESs are generally smooth functions.

Although the basis function types considered so far have performed quite satisfactory, it is interesting to consider other types and see how they perform. For example, one idea would be using localized basis functions like splines or wavelets. Along this direction, Demuynck et al. (2017) employed VES with one-dimensional bias potential written as a sum of Gaussians, though the performance of this choice was not compared to other basis function types.

## 2.5    Target Distribution

A strength of variationally enhanced sampling is its flexibility in choosing the target distribution $p(\mathbf{s})$, which allows to tailor the sampling of CVs. Generally one wants to employ a target distribution where the sampling is enhanced as compared to the unbiased distribution $P(\mathbf{s})$. The simplest choice is to consider a uniform target distribution, in other words try to completely flatten the sampling in CV space.

However, this is often rather non-optimal as it leads to sampling regions high in free energy that are generally not of interest. Instead, it is usually better to take inspiration from well-tempered metadynamics (Barducci et al. 2008) and consider the so-called well-tempered distribution given by

$$p(\mathbf{s}) = \frac{[P(\mathbf{s})]^{1/\gamma}}{\int d\mathbf{s}\,[P(\mathbf{s})]^{1/\gamma}} = \frac{e^{-(1/\gamma)\beta F(\mathbf{s})}}{\int d\mathbf{s}\,e^{-(1/\gamma)\beta F(\mathbf{s})}},$$

(13)

where $\gamma > 1$ is an adjustable parameter called the bias factor that controls how much the sampling is enhanced as compared to the unbiased distribution $P(\mathbf{s})$. In this way it is possible to enhance CV fluctuations in a controllable manner and focus the sampling toward the lowest free energy regions that are most relevant. This distribution can also be viewed as sampling on an effective free energy landscape where the barriers have been reduced by a factor of $\gamma$. Within VES the dependence on the a priori unknown $F(\mathbf{s})$ is bypassed by using a simple iterative scheme (Valsson and Parrinello 2015).

In general using the well-tempered distribution as target distribution is the most appropriate choice. However, in specific cases it can be useful to use other choices, for example, to localize the sampling on a specific region in CV space or to enforce restraints on the CVs.

## 2.6    Reweighting

From the bias potential it is possible to directly estimate the FES as a function of the biased CV using Eq. 6. However, it should be noted that by reweighting it is also possible to get the FES as a function of any other set of CVs or to get the unbiased expectation values of any observable that depends on the atomic coordinates.

From the relation $P(\mathbf{R}) \propto P_V(\mathbf{R})\,e^{\beta V(\mathbf{s}(\mathbf{R}))}$, one can write the unbiased average of an observable $O(\mathbf{R})$ as

$$\langle O(\mathbf{R}) \rangle = \frac{\left\langle O(\mathbf{R})\,e^{\beta V(\mathbf{s}(\mathbf{R}))} \right\rangle_V}{\left\langle e^{\beta V(\mathbf{s}(\mathbf{R}))} \right\rangle_V},$$

(14)

where the expectation values on the right-hand side are obtained in the biased ensemble. In other words, each configuration obtained in the biased ensemble is weighted by the value of the bias acting on it. It is possible to employ this reweighting even during the VES optimization if the bias potential is changing sufficiently slow that it can be considered as quasi-stationary.

The FES as a function of any set of CVs $\mathbf{s}'$ can be obtained by using the relation

$$F(\mathbf{s}') = -\frac{1}{\beta} \log \left\langle \delta(\mathbf{s}' - \mathbf{s}'(\mathbf{R})) \right\rangle = -\frac{1}{\beta} \log \left\langle \delta(\mathbf{s}' - \mathbf{s}'(\mathbf{R}))\,e^{\beta V(\mathbf{s}(\mathbf{R}))} \right\rangle_V,$$

(15)

which in practice is obtained by computing a reweighted histogram. It can happen that the form of the bias potential employed is efficient enough to enhance the sampling but might not have sufficient variational flexibility to fully represent the FES. In such cases one can use reweighting to obtain the FES for the biased CVs. In any case, it is a good general policy to obtain the FES by reweighting and compare that to the one obtained directly from the bias potential, significant difference between the two might indicate that there are some problems with the simulation, for instance, the CVs are poorly chosen or the basis set size is not sufficient to describe the FES.

## 2.7    Software Implementation

Variationally enhanced sampling is available through the VES code (http://www. ves-code.org) which is a module for the PLUMED 2 molecular dynamics plug-in (Tribello et al. 2014). This makes it possible to use VES with all the MD codes that can be interfaced with PLUMED 2 and make use of its vast library of CVs. The VES code is open-source and is officially integrated into PLUMED since version 2.4. We refer the reader to the manual of PLUMED 2 to see the features available in the VES code.

## 3    Applications and Extensions of Variationally Enhanced Sampling

Variationally enhanced sampling has already been used in various applications (Piccini et al. 2017; Piaggi et al. 2017; Palazzesi et al. 2017; Demuynck et al. 2017; Perego et al. 2018; Yang and Parrinello 2018). However, the most intriguing aspect of VES is that it is based on a variational principle. This can be used in novel and innovative ways by employing clever choices for the bias potential and the target distribution. In some sense VES can be viewed as a framework for developing new ideas and methods within enhanced sampling and free energy calculations. Already there are numerous interesting applications of VES that utilize the variational principle in this direction. We will briefly describe them in the following.

## 3.1    Free-Energy Flooding for Kinetic Rate Calculations

Obtaining kinetic properties of rare events from molecular simulations is of great practical interest, for example, how much time on average the system takes to go from one metastable state to another. Inspired by the success of using metadynamics to obtain kinetics of rare events from biased simulation (Tiwary and Parrinello 2013), an extension of VES has been developed for this purpose.

The main assumption is that the transition state should be free of bias. It can then be shown that following relations holds between the physical transition time $\tau$ and the transition time in the biased simulation $\tau_V$ (Voter 1997; Grubmüller 1995)

$$\tau = \tau_V \left\langle e^{\beta V(\mathbf{s})} \right\rangle_V, \tag{16}$$

where $\langle e^{\beta V(\mathbf{s})} \rangle_V$ is the so-called acceleration factor measured in the biased simulation that tells how much the time is boosted as compared to the unbiased simulation.

The idea that McCarty et al. (2015) followed is to use VES to generate a bias potential that only fills the underlying FES up to some predefined cutoff value $F_c$. Thus, by appropriately choosing the cutoff value, it is possible to significantly accelerate transitions between metastable states while still ensuring that the transition states are free of bias, thus allowing the usage of Eq. 16. This is achieved by considering a bias potential of the form

$$V(\mathbf{s}; \boldsymbol{\alpha}) = v(\mathbf{s}; \boldsymbol{\alpha}) \cdot S(-v(\mathbf{s}; \boldsymbol{\alpha}) - F_c), \tag{17}$$

where $v(\mathbf{s}; \boldsymbol{\alpha})$ is some representation of the FES (e.g., a linear expansion as in Eq. 11) and $S(x)$ is a sigmoidal switching function that continuously goes from 1 to 0 around $x = 0$. This is combined with a target distribution given by

$$p(\mathbf{s}) = \frac{S(F^*(\mathbf{s}) - F_c)}{\int d\mathbf{s}\, S(F^*(\mathbf{s}) - F_c)}, \tag{18}$$

that is iteratively updated using the current best estimate of the FES $F^*(\mathbf{s}) \approx -v(\mathbf{s})$.

In practical applications the bias potential is first optimized with a given cutoff value $F_c$ until a satisfactory convergence has been achieved. The bias potential is then fixed, and many independent simulations are performed to obtain the necessary statistics. It is a good practice to run at least two sets of simulations using different cutoff values in order to check the consistency of the results. The physical time in each simulation is simply obtained by rescaling the MD time step $\Delta t_{\mathrm{MD}}$ by the bias acting at each MD step

$$t_0 = \sum_i^{n_{\mathrm{MD}}} \Delta t_{\mathrm{MD}} \cdot e^{\beta V(\mathbf{s}(t_i))}, \tag{19}$$

where $n_{\mathrm{MD}}$ is the number of MD steps and $\mathbf{s}(t_i)$ is the value of the CVs at time $t_i = i\, \Delta t_{\mathrm{MD}}$.

The method has been used to obtain kinetic rates for chemical reactions (Piccini et al. 2017) and unfolding times of small proteins (Palazzesi et al. 2017).

### 3.2    Approximate Bias Potentials for Exploring High-Dimensional Free Energy Landscapes

It is of great importance to be able to efficiently sample high-dimensional free energy landscapes. Unfortunately, CV-based enhanced sampling methods are generally limited in the number of CVs that can be biased at the same time. Various ideas and methods have thus been introduced that tackle this problem (Piana and Laio 2007; Abrams and Vanden-Eijnden 2010; Pfaendtner and Bonomi 2015).

A rather interesting idea to approach this problem within VES is to employ approximate bias potentials. One of the simplest examples would be to consider a bias potential that is a sum of pairwise terms

$$V (s_1, s_2, \ldots, s_d; \boldsymbol{\alpha}) = \sum_{i,j} V(s_i, s_j; \boldsymbol{\alpha}^{(i,j)}), \tag{20}$$

where for each pairwise term we employ a linear expansion in some basis functions.

This idea was pursued by Shaffer et al. (2016a) which considered the folding of a small protein where the conformational landscape is described by the large number of backbone dihedral angles. They employed a bias potential written as sum of pairwise terms where only nearest-neighbor dihedral angles along the protein backbone were explicitly correlated. The authors then combined this with a clever choice of the target distribution that indirectly accounts for the important long-range correlations that are missing in the approximate bias potential. In this way the authors were able to generate a bias potential that allowed for an efficient sampling of the folding process. In a follow-up paper (Shaffer et al. 2016b), the authors further extended this idea and showed how it can be used to construct simple coarse-grained models for protein dynamics.

### 3.3    Bespoke Bias for Obtaining Free Energy Differences

An intriguing possibility that VES offers is to use purpose-built bias models that depend on a few variational parameters that are optimized. In this spirit, McCarty et al. (2016) considered a bespoke bias potential based on a model for a two-state free energy landscape given by

$$F(\mathbf{s}_A, \mathbf{s}_B) = \frac{F_A(\mathbf{s}_A) + F_B(\mathbf{s}_B)}{2} - \sqrt{\left(\frac{F_A(\mathbf{s}_A) - F_B(\mathbf{s}_B) - \Delta F}{2}\right)^2 + \Delta^2}, \tag{21}$$

where $F_A(\mathbf{s}_A)$ and $F_B(\mathbf{s}_B)$ are free energy models for the two metastable states, $\Delta F$ is a proxy for the free energy difference between the two states, and $\Delta$ is a coupling parameter. The free energy models used in this equation only need to be locally valid descriptions of the two metastable states, thereby bypassing the need to find

CVs that globally describe the full conformational space of interest. By optimizing the two parameters $\Delta F$ and $\Delta$, it is possible to obtain a bias potential that leads to an effective sampling of the two metastable states.

## 3.4    Bespoke Bias from Physical Models

Similar in spirit to the previous example, Piaggi et al. (2016) followed an interesting idea where they designed a bias potential based on a known physical model. In particular, they considered the important subject of nucleation that is rather challenging for molecular simulations due to the long time scale on which the nucleation process occurs. By employing a bias potential based on a free energy model taken from classical nucleation theory, they were able to greatly accelerate the sampling of nucleation events. This example shows clearly the value of employing a bespoke bias potential based on a physically motivated functional form within VES.

## 3.5    Coarse-Graining from Variationally Enhanced Sampling

A common way to tackle complex system is to construct mesoscopic coarse-grained models that are able to describe its main physical behavior. However, such models introduce phenomenological parameters that need to be derived from the microscopic system. In general this is not an easy task. Invernizzi et al. (2017) showed that VES can be turned into a powerful method to systematically obtain these phenomenological parameters from the microscopic simulations. In particular, they considered the Ginzburg-Landau model for a second-order critical point and showed how VES can be used to obtain its phenomenological parameters in the region close to the liquid-vapor critical point of a Lennard-Jones fluid. The procedure introduced by the authors is completely general and can be adapted for other coarse-grained models.

## 3.6    Variational Approach to Monte Carlo Renormalization Group

A rather interesting application of VES to critical phenomena was reported by Wu and Car (2017). They used VES to develop a method for Monte Carlo renormalization group theory where the critical slowing down near a critical point is avoided by using a bias potential. By optimizing the bias potential within VES, it is possible to obtain an accurate estimate of the renormalized coupling constants and the critical exponents. The authors applied their method to the two-dimensional Ising model. The method can be extended to tackle critical phenomena in more complex systems. For instance, the authors mention it should be possible to adapt the method to tackle field theoretical models, such as the Ginzburg-Landau model. Thus, by combining their method with the phenomenological coarse-graining method from the previous Sect. (3.5), it would be possible to investigate critical phenomena starting from the atomistic scale.

# 4    Conclusion and Outlook

Despite the relatively short time since its introduction, variationally enhanced sampling has already been used in numerous innovative and interesting applications, as we have discussed in the previous section. It is our firm belief that many more will come in the future. Furthermore, we believe that the performance of VES can be further improved in many ways. Thus, the future of VES looks bright and there are many new innovations to be expected.

**Index terms**  atomistic simulations; kinetic bottlenecks; metastable states; molecular dynamics; collective variables; free energy calculations; free energy landscapes; enhanced sampling; kinetic rate calculations; variational principle.

# References

Abrams CF, Vanden-Eijnden E (2010) Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. Proc Natl Acad Sci USA 107(11):4961–4966. https://doi.org/10.1073/pnas.0914540107

Bach F, Moulines E (2013) Non-strongly-convex smooth stochastic approximation with convergence rate o(1/n). In: Burges C, Bottou L, Welling M, Ghahramani Z, Weinberger K (eds) Advances in neural information processing systems, vol 26. Curran Associates, Inc., Red Hook, pp 773–781

Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett 100(2):020603

Bilionis I, Koutsourelakis P (2012) Free energy computations by minimization of Kullback–Leibler divergence: an efficient adaptive biasing potential method for sparse representations. J Comput Phys 231(9):3849–3870. https://doi.org/10.1016/j.jcp.2012.01.033

Bussi G, Gervasio FL, Laio A, Parrinello M (2006) Free-energy landscape for $\beta$ hairpin folding from combined parallel tempering and metadynamics. J Am Chem Soc 128(41):13435–13441. https://doi.org/10.1021/ja062463w

Darve E, Pohorille A (2001) Calculating free energies using average force. J Chem Phys 115(20):9169. https://doi.org/10.1063/1.1410978

Demuynck R, Rogge SMJ, Vanduyfhuys L, Wieme J, Waroquier M, Van Speybroeck V (2017) Efficient construction of free energy profiles of breathing metal–organic frameworks using advanced molecular dynamics simulations. J Chem Theory Comput 13(12):5861–5873. https://doi.org/10.1021/acs.jctc.7b01014

Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: conformational flooding. Phys Rev E 52(3):2893–2906. https://doi.org/10.1103/PhysRevE.52.2893

Hansmann U, Wille L (2002) Global optimization by energy landscape paving. Phys Rev Lett 88(6):068105. https://doi.org/10.1103/PhysRevLett.88.068105

Huber T, Torda AE, Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aided Mol Des 8(6):695–708. https://doi.org/10.1007/BF00124016

Invernizzi M, Valsson O, Parrinello M (2017) Coarse graining from variationally enhanced sampling applied to the Ginzburg–Landau model. Proc Natl Acad Sci 114(13):3370–3374. https://doi.org/10.1073/pnas.1618455114

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99(20):12562–12566. https://doi.org/10.1073/pnas.202427399

Maragakis P, van der Vaart A, Karplus M (2009) Gaussian-mixture umbrella sampling. J Phys Chem B 113(14):4664–4673. https://doi.org/10.1021/jp808381s

McCarty J, Valsson O, Tiwary P, Parrinello M (2015) Variationally optimized free-energy flooding for rate calculation. Phys Rev Lett 115(7). https://doi.org/10.1103/PhysRevLett.115.070601

McCarty J, Valsson O, Parrinello M (2016) Bespoke bias for obtaining free energy differences within variationally enhanced sampling. J Chem Theory Comput 12(5):2162–2169. https://doi.org/10.1021/acs.jctc.6b00125

Palazzesi F, Valsson O, Parrinello M (2017) Conformational entropy as collective variable for proteins. J Phys Chem Lett 8(19):4752–4756. https://doi.org/10.1021/acs.jpclett.7b01770

Perego C, Valsson O, Parrinello M (2018) Chemical potential calculations in non-homogeneous liquids. J Chem Phys 149(7):072305. https://doi.org/10.1063/1.5024631

Pfaendtner J, Bonomi M (2015) Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. J Chem Theory Comput 11(11):5062–5067. https://doi.org/10.1021/acs.jctc.5b00846

Piaggi PM, Valsson O, Parrinello M (2016) A variational approach to nucleation simulation. Faraday Discuss 195:557–568. https://doi.org/10.1039/c6fd00127k

Piaggi PM, Valsson O, Parrinello M (2017) Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. Phys Rev Lett 119(1). https://doi.org/10.1103/PhysRevLett.119.015701

Piana S, Laio A (2007) A bias-exchange approach to protein folding. J Phys Chem B 111(17):4553–4559. https://doi.org/10.1021/jp067873l

Piccini G, McCarty JJ, Valsson O, Parrinello M (2017) Variational flooding study of a SN2 reaction. J Phys Chem Lett 8(3):580–583. https://doi.org/10.1021/acs.jpclett.6b02852

Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. J Phys Chem B 110(8):3533–3539. https://doi.org/10.1021/jp054359r

Rubinstein R (1999) Methodology and computing in applied probability. 1(2):127–190. https://doi.org/10.1023/A:1010091220143

Shaffer P, Valsson O, Parrinello M (2016a) Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin. Proc Natl Acad Sci 113(5):1150–1155. https://doi.org/10.1073/pnas.1519712113

Shaffer P, Valsson O, Parrinello M (2016b) Hierarchical protein free energy landscapes from variationally enhanced sampling. J Chem Theory Comput 12(12):5751–5757. https://doi.org/10.1021/acs.jctc.6b00786

Shell MS (2008) The relative entropy is fundamental to multiscale and inverse thermodynamic problems. J Chem Phys 129(14):144108. https://doi.org/10.1063/1.2992060

Tiwary P, Parrinello M (2013) From metadynamics to dynamics. Phys Rev Lett 111(23):230602. https://doi.org/10.1103/PhysRevLett.111.230602

Torrie G, Valleau J (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. J Comput Phys 23(2):187–199. https://doi.org/10.1016/0021-9991(77)90121-8

Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) Plumed 2: new feathers for an old bird. Comput Phys Commun 185(2):604–613. https://doi.org/10.1016/j.cpc.2013.09.018

Valsson O, Parrinello M (2014) Variational approach to enhanced sampling and free energy calculations. Phys Rev Lett 113(9):090601. https://doi.org/10.1103/PhysRevLett.113.090601

Valsson O, Parrinello M (2015) Well-tempered variational approach to enhanced sampling. J Chem Theory Comput 11(5):1996–2002

Valsson O, Tiwary P, Parrinello M (2016) Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. Annu Rev Phys Chem 67(1):159–184. https://doi.org/10.1146/annurev-physchem-040215-112229

Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett 78(20):3908–3911. https://doi.org/10.1103/PhysRevLett.78.3908

Wang F, Landau D (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. Phys Rev Lett 86(10):2050–2053. https://doi.org/10.1103/PhysRevLett.86.2050

Wu Y, Car R (2017) Variational approach to Monte Carlo renormalization group. Phys Rev Lett 119(22). https://doi.org/10.1103/PhysRevLett.119.220602

Yang YI, Parrinello M (2018) Refining collective coordinates and improving free energy representation in variational enhanced sampling. J Chem Theory Comput. 14:2889–2894. https://doi.org/10.1021/acs.jctc.8b00231

Zhang W, Wang H, Hartmann C, Weber M, Schütte C (2014) Applications of the cross-entropy method to importance sampling and optimal control of diffusions. SIAM J Sci Comput 36(6):A2654–A2672. https://doi.org/10.1137/14096493X

# Water: Many-Body Potential from First Principles (From the Gas to the Liquid Phase)

**30**

Francesco Paesani

## Contents

**Abstract**

Computer simulations have become an integral part of the toolbox of any researcher interested in molecular sciences, often providing new insights that are difficult (if not impossible) to obtain by other means. However, the predictive power of a computer simulation directly depends on the level of realism that can be used to represent the molecular system of interest. Since the early times of computer simulations, the search for a molecular model of water capable of describing its unique properties across different phases has been the focus of intense research. The continued increase in computer power accompanied by advances in the design of efficient algorithms for correlated electronic structure calculations and tremendous progress in the representation of global potential energy surfaces have recently opened the doors to the development of molecular models rigorously derived from many-body expansions of interaction energies. By quantitatively reproducing individual interaction terms between molecules, it has been shown that these many-body potential energy functions

F. Paesani (✉)
Department of Chemistry and Biochemistry, Materials Science and Engineering, and San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA
e-mail: fpaesani@ucsd.edu

635

can achieve unprecedented accuracy in computer simulations. This chapter provides an overview of the theoretical formalism underlying such many-body representations, with a particular focus on the performance of the MB-pol potential energy function in predicting the energetics as well as structural, thermodynamic, dynamical, and spectroscopic properties of water from the gas to the condensed phase.

# 1 Introduction

The importance of water cannot be overemphasized. This is the case primarily because the combination of unique physical properties and the wide range of conditions in which it exists in the liquid phase make water essential to life as we know it (Maréchal 2007). After decades of research, there is now little question that the unique behavior of water does not simply result from a three-atom, 10-electron molecule, but rather from the delicate interplay of many-body interactions between individual water molecules, which is further modulated by temperature and pressure (Eisenberg and Kauzmann 1969).

Given water's key role as life's matrix (Ball 2008), it is not surprising that many theoretical and computational studies have attempted to derive a molecular-level picture of water from the gas to the condensed phase. Since the first Monte Carlo (MC) and molecular dynamics (MD) simulations of liquid water performed respectively by Barker and Watts (1969), and Rahman and Stillinger (1971), a myriad of models have been developed, based either on molecular mechanics (commonly defined as "force fields") or quantum mechanics (commonly defined as "ab initio" models). Various flavors of force fields, with different degrees of empiricism, have been proposed over the years, ranging from coarse-grained representations with no atomistic details to classical parameterizations in terms of point charges and rigid bonds, and more sophisticated models that account for molecular flexibility, electronic polarization, and charge transfer (Guillot 2002; Vega and Abascal 2011; Cisneros et al. 2016; Shvab and Sadus 2016). Despite much progress, it has been shown that even the most successful and popular water force fields still exhibit limited accuracy and effectively lack predictive power (Cisneros et al. 2016).

On the other hand, ab initio models are based on either wavefunction theory (WFT) or density functional theory (DFT). Although correlated WFT methods, such as coupled cluster with single, double, and perturbative triple excitations, commonly abbreviated as CCSD(T), can enable molecular level studies of water with chemical accuracy, usually defined as a deviation of 1 kcal/mol from the "exact" result (Rezac and Hobza 2013; Simova et al. 2013), the associated computational cost limits their application to systems containing no more than a handful of molecules (Manna et al. 2017). This implies that DFT remains the only viable ab initio approach for simulations of water across different phases. However, existing DFT models are affected by several shortcomings that drastically limit their predictive ability in

describing the properties of water from small clusters in the gas phase to the liquid phase and ice (Gillan et al. 2016).

As an alternative to both force fields and ab initio models, a new class of analytical potential energy functions (PEFs) has recently been introduced. These PEFs are derived from a rigorous representation of the many-body expansion (MBE) of the interaction energy ($V_N$) between N water molecules (Hankins et al. 1970)

$$V_N (r_1, \ldots, r_N) = \sum_{i=1}^{N} V_{1B} (r_i) + \sum_{i<j}^{N} V_{2B} (r_i, r_j)$$
$$+ \sum_{i<j<k}^{N} V_{3B} (r_i, r_j, r_k) + \cdots + V_{NB} (r_1, r_2, r_3, \ldots, r_N) \tag{1}$$

where $r_i$ collectively denotes the coordinates of the oxygen and hydrogen atoms within the $i^{th}$ water molecule. In Eq. (1), $V_{1B}$ is the one-body (1B) potential, representing the energy required to deform an individual water molecule from its equilibrium geometry, while all higher n-body (nB) interaction terms ($V_{nB}$) are defined recursively through

$$V_{nB} (r_1, \ldots, r_n) = V_n (r_1, \ldots, r_n) - \sum_{i=1}^{N} V_{1B} (r_i) - \sum_{i<j}^{N} V_{2B} (r_i, r_j) - \ldots$$
$$- \sum_{i<j<k<\cdots<n-1}^{N} V_{(n-1)B} (r_i, r_j, r_k, \ldots, r_{n-1}). \tag{2}$$

The importance of many-body effects in water was already recognized in the 1950s by Frank and Wen who introduced a molecular picture of liquid water consisting of "flickering clusters of hydrogen-bonded molecules," emphasizing the cooperative nature of hydrogen bonding (Frank and Wen 1957). The first attempts to develop many-body (MB) potential energy functions (PEFs) for water only using ab initio data were made by Clementi and co-workers (Matsuoka et al. 1976; Evans et al. 1987; Niesar et al. 1990). These pioneering studies demonstrated that Eq. (1) converges rapidly for water, showing that 3B effects can contribute as much as 15–20% to the interaction energy of cyclic water structures and 4B effects represent, on average, 1% of the interaction energy.

With the continued development of faster hardware and efficient algorithms for correlated WFT methods, it is nowadays routine to perform calculations of accurate interaction energies for small molecular systems, which can then serve as a reference for deriving analytical representations of the individual terms of Eq. (1). In particular, exploiting the rapid convergence of the MBE for water, the "Stratified Approximation Many-Body Approach" (SAMBA) has been shown

to provide highly accurate interaction energies for water clusters through the application of progressively lower-level electronic structure methods to represent subsequently higher-body terms of the MBE (Gora et al. 2011). In parallel with hardware and algorithmic developments, tremendous progress has been made in constructing multidimensional mathematical functions that can accurately represent global potential energy surfaces with a large number of degrees of freedom (Braams and Bowman 2009; Bartók et al. 2010; Behler 2016). This progress has been accompanied by the development of efficient classical schemes for treating electrostatic interactions in many-body molecular systems (Thole 1981; van Duijnen and Swart 1998; Cisneros et al. 2006; Piquemal et al. 2006; Torheyden and Jansen 2006; Cisneros et al. 2007; Handley et al. 2009; Gao et al. 2014). As a result, several MB PEFs have been proposed in the last years, the most notable of which are CC-pol (Bukowski et al. 2007), WHBB (Wang et al. 2011) HBB2-pol (Babin et al. 2012), and MB-pol (Babin et al. 2013, 2014; Medders et al. 2014).

Among existing many-body PEFs, MB-pol has been shown to correctly predict the properties of water across different phases (Reddy et al. 2016), reproducing the vibration-rotation tunneling spectrum of the water dimer (Babin et al. 2013), the energetics, quantum equilibria, and infrared spectra of small clusters (Babin et al. 2014; Brown et al. 2017), the structural, thermodynamic, and dynamical properties of liquid water (Medders et al. 2014), the energetics of ice phases (Pham et al. 2017), the infrared (IR) and Raman spectra of liquid water (Medders and Paesani 2015; Reddy et al. 2017), the vibrational sum-frequency generation spectrum of the air/water interface (Medders and Paesani 2016), and the IR and Raman spectra of ice $I_h$ (Moberg et al. 2017).

This chapter reviews recent progress in modeling the properties of water across different phases using computer simulations with the MB-pol PEF. After a brief overview of the theoretical formalism underlying MB-pol and associated computational algorithms, a systematic analysis of the MBE for water is presented, with comparisons with both force fields and ab initio models. The accuracy of MB-pol in representing the multidimensional Born-Oppenheimer potential energy surface of water in the liquid phase and ice is then assessed. Finally, the ability of MB-pol to predict structural, thermodynamic, and dynamical properties as well as vibrational spectra of water from small clusters in the gas phase to the liquid phase and ice is discussed. This is followed by a brief outlook on future developments and applications of many-body PEFs in computer simulations of generic molecular systems.

## 2    MB-pol Theoretical Framework and Functional Form

MB-pol was rigorously derived from Eq. (1). It includes explicit 1B, 2B, and 3B terms, while all higher-order contributions are accounted for in a mean field sense through a classical representation of N-body induction (Babin et al. 2013, 2014). Each water molecule is represented by six sites: three sites correspond to the physical atoms (oxygen and two hydrogen atoms), two fictitious sites are located

**Fig. 1** MB-pol model of a water molecule. O, $H_1$, and $H_2$ correspond to the oxygen and hydrogen atoms, respectively, $L_1$ and $L_2$ represent the oxygen lone-pairs, and M is an additional fictitious site
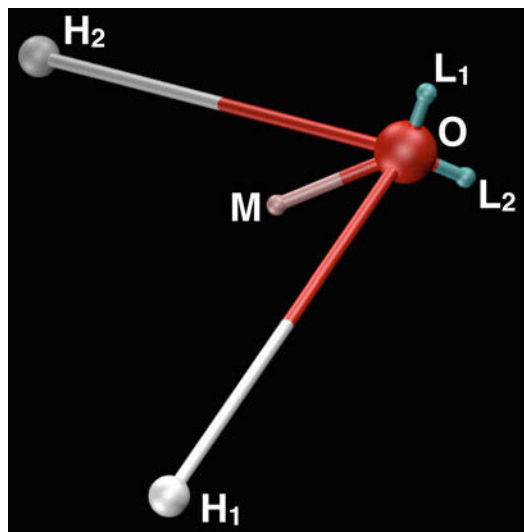
symmetrically along the directions of the oxygen lone-pairs, and an additional M-site is located along the HOH bisector (Fig. 1). The number and location of the six sites were optimized to reproduce both dipole and quadrupole moments of an isolated water molecule as well as to accurately represent the anisotropy of the potential energy surface of the water dimer.

The 1B term of MB-pol corresponds to the monomer PEF developed by Partridge and Schwenke which accurately reproduces the rovibrational spectrum of an isolated water molecule in the gas phase (Partridge and Schwenke 1997). Associated with the 1B energy expression is a nonlinear dipole moment surface represented by geometry dependent (positive) point charges on the two hydrogen atoms, which are balanced by a negative point charge placed on the M-site (Fig. 1). The molecular polarizabity tensor of an isolated water molecule is represented by isotropic atomic polarizabilities located on the three physical atoms.

MB-pol 2B term is expressed by combining classical expressions for permanent and induced electrostatics, and dispersion energy with a multidimensional term that corrects for deficiencies associated with a purely classical representation of intermolecular interactions at short range (Stone 1997). This additional term, which smoothly switches to zero as the separation between the oxygen atoms of the two water molecules within a dimer approaches 6.5 Å, effectively represents quantum-mechanical interactions arising from the overlap of the monomer electron densities (Babin et al. 2013), such as charge transfer and penetration and Pauli repulsion (Khaliullin et al. 2007, 2008). Both 2B permanent electrostatics and induction are represented by a modified Thole-type expression derived from the TTM4-F model (Burnham et al. 2008), while the dispersion energy is expressed through a rigorous fit to the asymptotic ab initio reference energy as originally introduced by the CC-pol PEF (Bukowski et al. 2008). Following the same theoretical scheme,

MB-pol represents 3B interactions by combining a classical 3B induction term with a multidimensional short-range function, which smoothly switches to zero when the oxygen-oxygen separations between all pairs of water molecules within a trimer approach 4.5 Å (Babin et al. 2014). Given its functional form, MB-pol can then be viewed as a many-body PEF built upon a baseline polarizable model supplemented by short-range 2B and 3B contributions that effectively represent quantum-mechanical interactions that cannot be described by purely classical expressions (Medders and Paesani 2015).

The short-range 2B and 3B multidimensional functions are expressed by 4th-degree permutationally invariant polynomials (Braams and Bowman 2009) in variables that are functions of the distances between all six molecular sites (Fig. 1). The invariance is imposed with respect to permutations of whole molecules as well as of the two hydrogen atoms within the same molecule. The coefficients of both 2B and 3B permutationally invariant polynomials were optimized using supervised learning (Abu-Mostafa et al. 2012) to reproduce the interaction energies calculated at the CCSD(T) level of theory in the complete basis set (CBS) limit for approximately 42,000 and 12,000 dimers and trimers, respectively. Specific details about MB-pol 2B and 3B training sets can be found in the original references (Babin et al. 2013, 2014).

As shown in Fig. 2, MB-pol provides an accurate description of the low-order terms (i.e., 2B, 3B, and 4B) of the MBE. Specifically, the 2B term is associated with a root-mean-square-deviation (RMSD) of 0.028 kcal/mol relative to the CCSD(T)/CBS reference values. The latter were obtained via a two-point extrapolation (Halkier et al. 1999a, b) of the energies calculated with aug-cc-pVTZ and aug-cc-pVQZ basis sets (Dunning 1989) supplemented with an additional set of midbond (mb) functions (Tao and Pan 1992) placed at the center of mass of each dimer configuration. The RMSD associated with the MB-pol 3B term is 0.026 kcal/mol relative to CCSD(T)/CBS reference values obtained with the aug-cc-pVTZ basis set supplemented with the same set of midbond functions used for the 2B calculations, which were placed at the center of mass of each trimer configuration. Finally, the RMSD for the 4B interactions is 0.030 kcal/mol relative to MP2 reference values obtained with the aug-cc-pVTZ basis set supplemented with the same set of midbond functions described above, which were placed at the center of mass of each tetramer.

The accuracy of MB-pol in reproducing the individual terms of the MBE (Fig. 2) reflects the ability of the underlying functional form to correctly represent many-body effects at both short and long intermolecular separations. To further demonstrate this aspect, Fig. 3 shows a comparison between the results obtained with MB-pol and the absolutely-localized molecular orbital energy decomposition analysis, ALMO EDA, method (Khaliullin et al. 2007, 2008) for a radial scan along the hydrogen bond coordinate of the minimum energy structure of the water dimer. In this analysis, the O-O distance ($R_{OO}$) is constrained at different values while all other atomic coordinates are optimized. Panel (a) shows the comparison between MB-pol (Thole-damped) permanent and induced electrostatics and the corresponding terms obtained from ALMO EDA calculations carried out at the
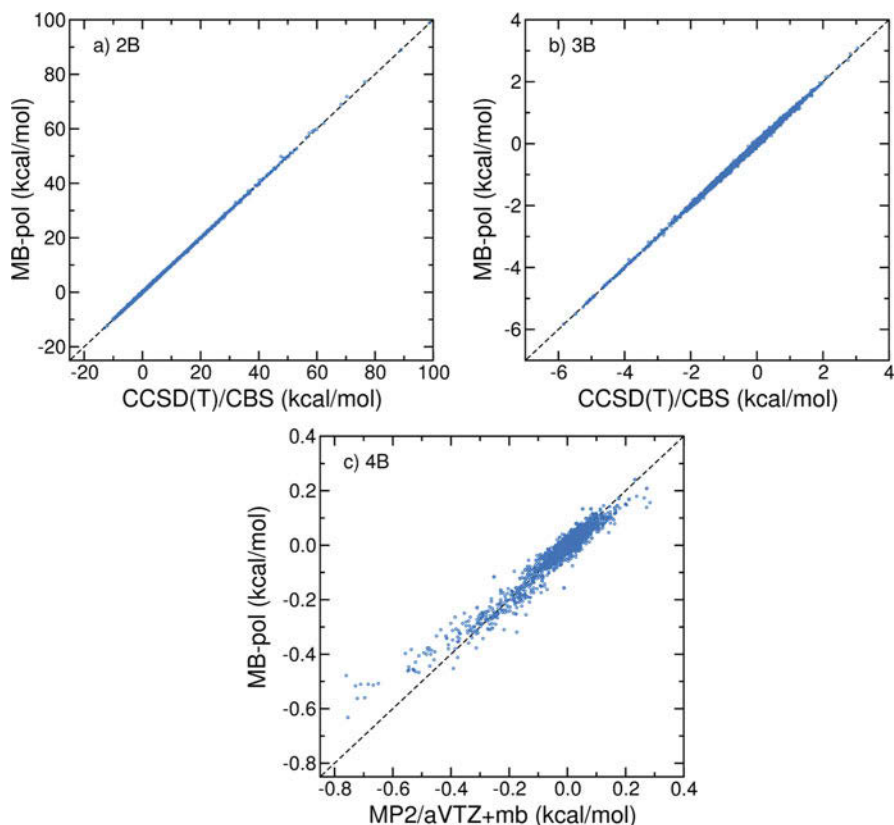
**Fig. 2** Correlation plots of 2B (**a**), 3B (**b**), and 4B (**c**) interaction energies. On the x-axes are the reference values calculated at the CCSD(T)/CBS (for 2B and 3B) and MP2/aVTZ (for 4B) levels of theory. On the y-axes are the corresponding MB-pol values. (Adapted from Paesani 2016)

ωB97M-V level of theory (Mardirossian and Head-Gordon 2015) with the def2-TZVPPD basis set (Weigend and Ahlrichs 2005; Rappoport and Furche 2010). In this analysis, ωB97M-V was specifically chosen because it yields interaction energies in close agreement with the CCSD(T) reference data (see Fig. 5), while the def2-TZVPPD basis set is large enough to fully capture charge transfer but still relatively small to avoid inter-monomer-orbital linear-dependence.

The agreement between MB-pol and ALMO EDA for $R_{OO} > 3.5$ Å demonstrates that the baseline electrostatic model adopted by MB-pol correctly reproduces permanent and induced electrostatic interactions at medium and long range, where there is minimal, if any, overlap between the monomer electron densities. As expected, deviations due to charge penetration, which cannot be represented by classical expressions, are found at short range, particularly, for the contribution associated with permanent electrostatics. Similarly, panel (b) shows that the (damped) dispersion energy term of MB-pol quantitatively reproduces the corresponding
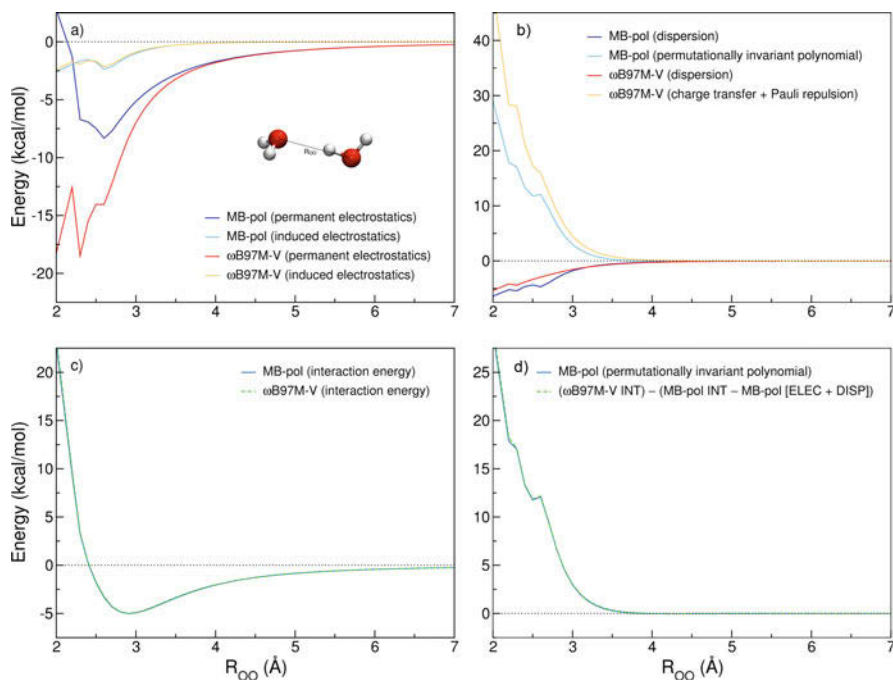
**Fig. 3** Comparison between MB-pol and ALMO EDA energy decomposition for a radial scan along the hydrogen bond coordinate of the minimum energy structure of the water dimer. (**a**) Comparison between permanent and induced electrostatic interactions obtained from MB-pol and ωB97M-V/def2-TZVPPD. (**b**) Comparison of the terms associated with dispersion energy and charge transfer combined with Pauli repulsion obtained from ωB97M-V/def2-TZVPPD calculations with the corresponding MB-pol terms. (**c**) Comparison of MB-pol and ωB97M-V/def2-TZVPPD total interaction energy. (**d**) Comparison between the nonclassical terms obtained with ωB97M-V/def2-TZVPPD and MB-pol permutationally invariant polynomial

ALMO EDA values at medium and long range, with deviations at short distances, where the definition of dispersion energy derived from the asymptotic expansion of the interaction energy breaks down (Stone 1997). The comparisons shown in panel (b) also demonstrate that the energy contribution associated with the MB-pol 2B permutationally invariant polynomial displays the same radial dependence as the combined charge transfer and Pauli repulsion terms calculated from ALMO EDA. However, as expected, there are noticeable differences between these two terms at short range since MB-pol's permutationally invariant polynomial must also recover effects associated with charge penetration that are not taken into account by the baseline classical electrostatic model (panel a).

After adding all MB-pol terms together, the comparison in panel (c) demonstrates that MB-pol quantitatively reproduces the total interaction energy predicted by ωB97M-V. Finally, panel (d) shows that when MB-pol contributions associated with electrostatic interactions and dispersion energy are subtracted from the ωB97M-V

total interaction energy, MB-pol's permutationally invariant polynomial quantitatively recovers all remaining terms (i.e., charge transfer and penetration, and Pauli repulsion) that, by construction, are not included in the classical baseline model adopted by MB-pol. The analysis presented in Fig. 2 thus provides evidence that, within the MB-pol functional form, the 2B and 3B permutationally invariant polynomials quantitatively represent quantum-mechanical effects that cannot be described by purely classical expressions.

As a final note, it should be emphasized that maintaining the same functional form introduced with the original implementation of MB-pol (Babin et al. 2013, 2014), higher accuracy can be achieved by increasing the order of the permutationally invariant polynomials used to represent short-range 2B and 3B interactions. The choice to limit the highest order to the 4th degree was dictated by the necessity of achieving the optimal compromise between accuracy and computational efficiency. It should also be noted that the adoption of permutationally invariant polynomials is simply one among several options that exist to accurately represent multidimensional potential energy surfaces. In this context, it has recently been shown (Thuong et al. 2018) that it is possible to achieve similar accuracy as the original MB-pol implementation by replacing the 2B and 3B permutationally invariant polynomials with analogous neural networks (Behler 2016) and Gaussian approximation potentials (Bartók et al. 2010).

## 3    Energetics of Water Clusters, Liquid Water, and Ice

The accuracy and transferability of MB-pol can be systematically assessed from the analysis of both energetics and many-body effects in small water clusters, liquid water, and ice. To this purpose, MB-pol interaction energies for the low-lying isomers of the tetramer, pentamer, and hexamer clusters are compared in Fig. 4 with the corresponding reference values calculated at the CCSD(T)-F12/VTZ-F12 level of theory (Adler et al. 2007; Peterson et al. 2008; Knizia et al. 2009) for MP2 optimized geometries (Bates and Tschumper 2009; Temelso et al. 2011). Since the interaction energies are defined as the cluster energies minus the energies of the individual molecules in the same geometry as in the cluster, they are not affected by differences in the 1B energies predicted by CCSD(T)-F12/VTZ-F12 and MB-pol. For all isomers of the different clusters, MB-pol provides interaction energies in close agreement with the reference values, with the differences always being within chemical accuracy (Medders and Paesani 2015; Paesani 2016).

Among small water clusters, the hexamer holds a special place since its low-lying isomers display three-dimensional structures reminiscent of hydrogen-bond arrangements found in liquid water and ice. Furthermore, due to small differences in the interaction energies between different isomers, the hexamer cluster generally serves as a prototypical system to assess the accuracy of force fields and ab initio models of water. Figure 5 shows comparisons between the interaction energies of the low-lying hexamer isomers calculated using several advanced polarizable force fields (Fig. 5a) and common DFT models, without and with dispersion contributions
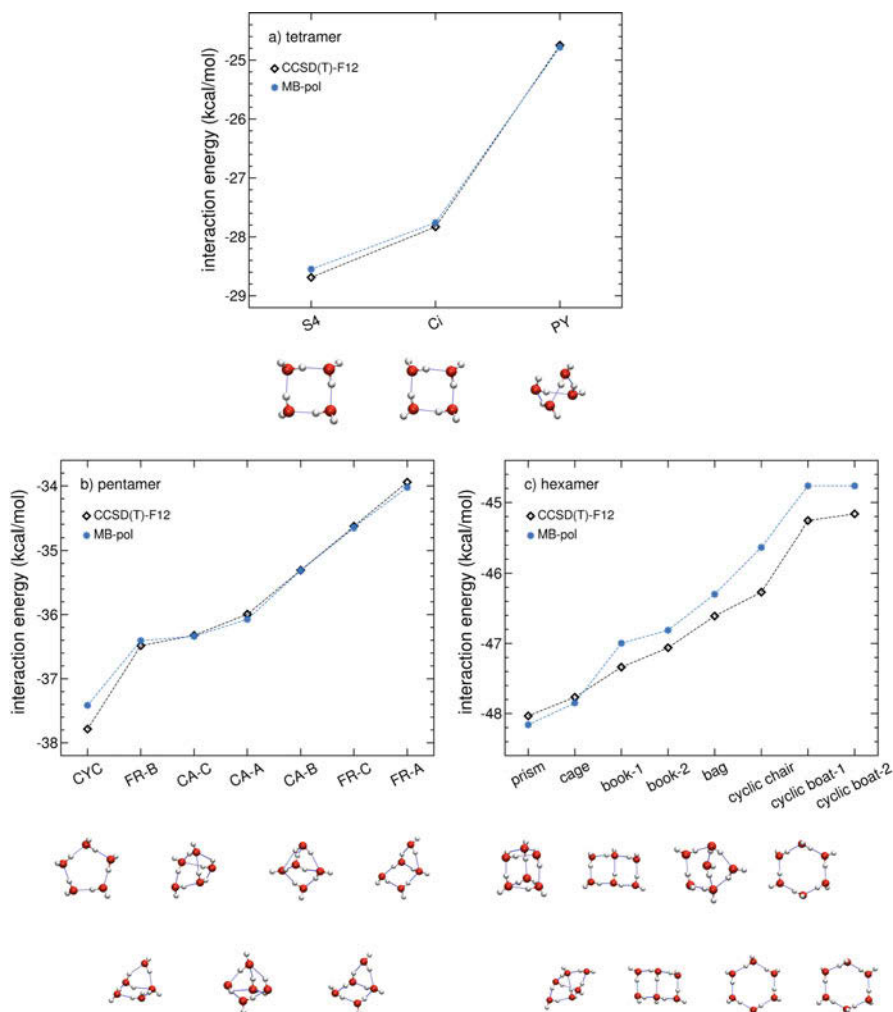
**Fig. 4** Comparisons between CCSD(T)-F12/VTZ-F12 and MB-pol interaction energies for the low-lying isomers of the tetramer (**a**), pentamer (**b**), and hexamer (**c**) clusters. The clusters geometries, shown at the bottom of the corresponding figures, were optimized at the MP2/avtz level of theory (Bates and Tschumper 2009; Temelso et al. 2011)

(Fig. 5b, c, respectively), and the corresponding CCSD(T)-F12/VTZ-F12 and MB-pol values.

While all polarizable force fields predict interaction energies relatively close to the reference values, only TTM3-F (Fanourgakis and Xantheas 2008) and WHBB (Wang et al. 2011) correctly reproduce the energy order. However, while the difference between TTM3-F and CCSD(T)-F12/VTZ-F12 is approximately of the

**Fig. 5** Interaction energies of the low-lying hexamer isomers calculated with advanced polarizable force fields (**a**), and DFT models without (**b**) and with (**c**) dispersion corrections. Also shown are the reference CCSD(T)-F12 and MB-pol values for comparison. Due to large variations in the interactions energies calculated with polarizable force fields and DFT models, different energy ranges are used in the three panels
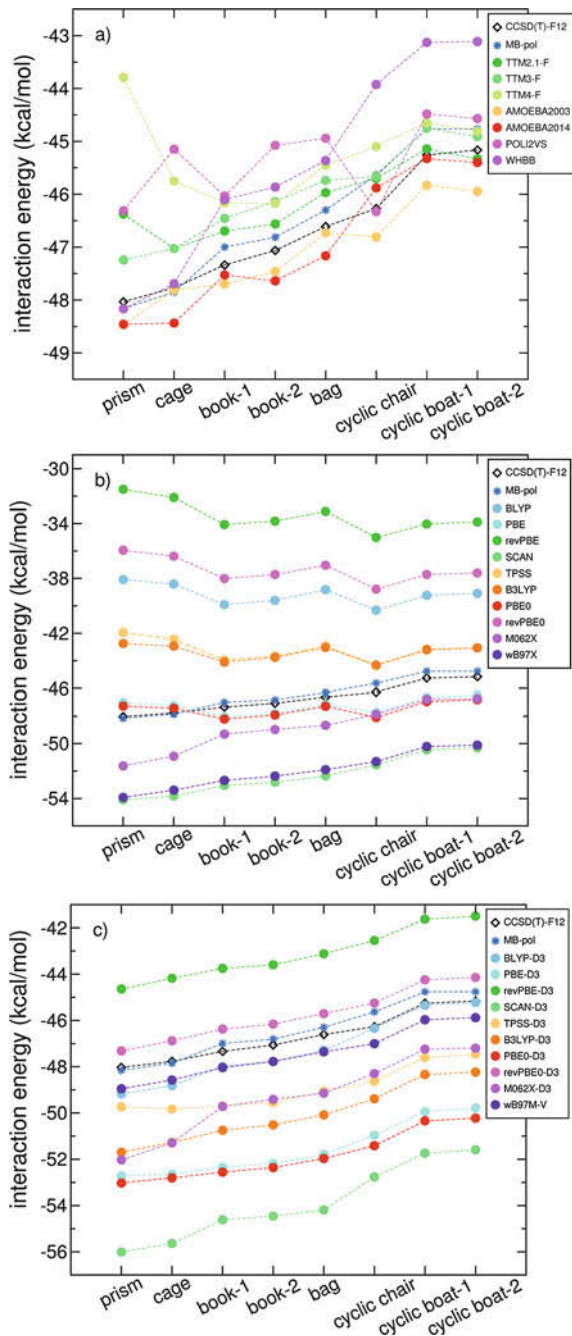
same magnitude for all isomers, WHBB, which, as MB-pol, was systematically derived from the MBE in Eq. (1), accurately describes the first two isomers (prism and cage) but predicts interaction energies that progressively deviate (up to 2 kcal/mol) from the reference CCSD(T)-F12/VTZ-F12 values for all other isomers. POLI2VS, which was developed to model vibrational spectra of water (Hasegawa and Tanimura 2011), completely fails to reproduce the energy progression from the prism to the cyclic boat-2 isomers and predicts the cyclic chair isomer, which lies ∼2 kcal/mol above the prism isomer at the CCSD(T)-F12/VTZ-F12 level of theory, to be the most stable isomer. Both versions of the AMOEBA force field, AMOEBA2003 (Ren and Ponder 2003) and AMOEBA2014 (Wang et al. 2013), qualitatively reproduce the overall trend, although the stability of some isomers is inverted. Interestingly, the TTM4-F model (Burnham et al. 2008) fails, by a large margin, to describe the energetics of the first three isomers, predicting prism to be the most unstable structure, while correctly reproduces the energy order of the higher-energy isomers.

Figure 5b, c shows analogous comparisons for several DFT models that are commonly used in computer simulations of water. Specifically, this analysis includes GGA functionals: BLYP (Becke 1988), PBE (Perdew et al. 1996), and revPBE (Zhang and Yang 1998); meta-GGA functionals: TPSS (Tao et al. 2003), and SCAN (Sun et al. 2015); hybrid GGA functionals: B3LYP (Becke 1993), PBE0 (Adamo and Barone 1999), and revPBE0 (Goerigk and Grimme 2011); meta hybrid GGA functionals: M06-2X (Zhao and Truhlar 2008); range-separated hybrid functionals: ωB97X (Chai and Head-Gordon 2008); and range-separated meta hybrid functionals: ωB97M-V (Mardirossian and Head-Gordon 2015). DFT energies were calculated without and with the empirical D3(0) dispersion correction (Grimme et al. 2010) except for ωB97M-V that, by construction, accounts for dispersion energy through the VV10 nonlocal functional (Vydrov and Van Voorhis 2010). All DFT calculations were carried out with the aug-cc-pVQZ basis set using Gaussian 09 (Frisch et al. 2009) except for revPBE, revPBE0, SCAN, and ωB97M-V for which Q-Chem 5.0 was used (Shao et al. 2015).

Without dispersion corrections, only M062X, ωB97X, and SCAN predict the correct energy order of the hexamer isomers, although the differences from the CCSD(T)-F12/VTZ-F12 reference values can be as large as ∼6 kcal/mol. Although the inclusion of dispersion energy contributions significantly improves the agreement with the reference data, none of the functionals examined in this analysis, with the exception of ωB97M-V, achieves the same accuracy as MB-pol.

To further investigate the ability of MB-pol to reproduce many-body effects in water, the interaction energies of the prism, cage, and cyclic chair isomers are decomposed into individual many-body contributions, and the errors relative to the CCSD(T)-F12/VTZ-F12 reference values $\left( \Delta E = \Delta E_{nB}^{model} - \Delta E_{nB}^{CCSD(T)} \right)$ are shown in Fig. 6. To put the MB-pol results in perspective, the same analysis is also carried out with the polarizable force fields and DFT models used in Fig. 5. Independently of the isomer, MB-pol reproduces, nearly quantitatively, the reference values for all terms of the MBE. Importantly, since, within MB-pol all terms higher than 3B are represented entirely by classical many-body induction, the

agreement between MB-pol and the reference values for these terms demonstrates that higher-order interactions in water are primarily electrostatic in nature.

Among functionals without dispersion corrections, PBE and PBE0 provide the closest agreement with the CCSD(T)-F12/VTZ-F12 values. However, both functionals are associated with appreciable errors at the 4B level. Interestingly, the corresponding modified versions (revPBE and revPBE0) largely underestimate 2B contributions, while providing similar accuracy for higher-body terms. The correct energy order predicted by M06-2X in Fig. 5 effectively results from (well-balanced) error cancellation between different interaction terms, which, individually, are associated with errors on the order of 1 kcal/mol even at the 5B level. Large 2B errors are also associated with BLYP, B3LYP, TPSS, SCAN, and ωB97X, with the first three functionals underestimating and the last two functionals overestimating the strength of the interactions in the water dimers, respectively. Explicit inclusion of dispersion contributions significantly improves the accuracy of the 2B terms predicted by BLYP, TPSS, B3LYP, revPBE, and revPBE0, but, at the same time, the 2B terms of PBE-D3, PBE0-D3, and SCAN-D3 become less accurate than those from the bare functionals. Overall, the best performance is provided by ωB97M-V, which effectively exhibits similar accuracy as MB-pol for all interaction terms.

An important aspect that emerges from the analysis reported in Fig. 6 is the inability of some functionals to correctly reproduce high-body interaction terms. As mentioned above, MB-pol is able to correctly reproduce all terms higher than 3B entirely through classical many-body induction. This implies that several functionals that are currently used in water simulations (e.g., PBE-D3, PBE0-D3, revPBE-D3, and revPBE0-D3) effectively provide a less accurate description of many-body interactions than a purely classical electrostatic model. Since, by construction, classical electrostatics relies on a multipole representation of the electron density, functionals that exhibit large errors for high-body interactions are likely unable to correctly describe the electron density (and associated electrostatic properties) of the water molecules.

After investigating water clusters, the final step to assess the ability of MB-pol to accurately describe the multidimensional Born-Oppenheimer surface of water across different phases is represented by the analysis of the energetics of liquid water and different ice phases. As discussed in the Introduction, due to the associated computational cost, CCSD(T) calculations are currently unaffordable for condensed phase systems, and quantum Monte Carlo (QMC) effectively becomes the reference method for molecular systems in periodic boundary conditions. Direct comparison with QMC relative energies calculated for molecular configurations extracted from simulations carried out with the vdW-DF and vdW-DF2 functionals (Morales et al. 2014) determined that MB-pol predicts the energetics of liquid water in close agreement with the reference values, resulting in a mean absolute deviation of 0.048 kcal/mol per molecule, which is more than two times smaller than the most accurate value obtained with popular DFT models commonly used in computer simulations of liquid water (Medders and Paesani 2015).

More recently, the energetics of several ice phases was also investigated with MB-pol (Pham et al. 2017). As reported in Table 1, in all cases, the lattice energies
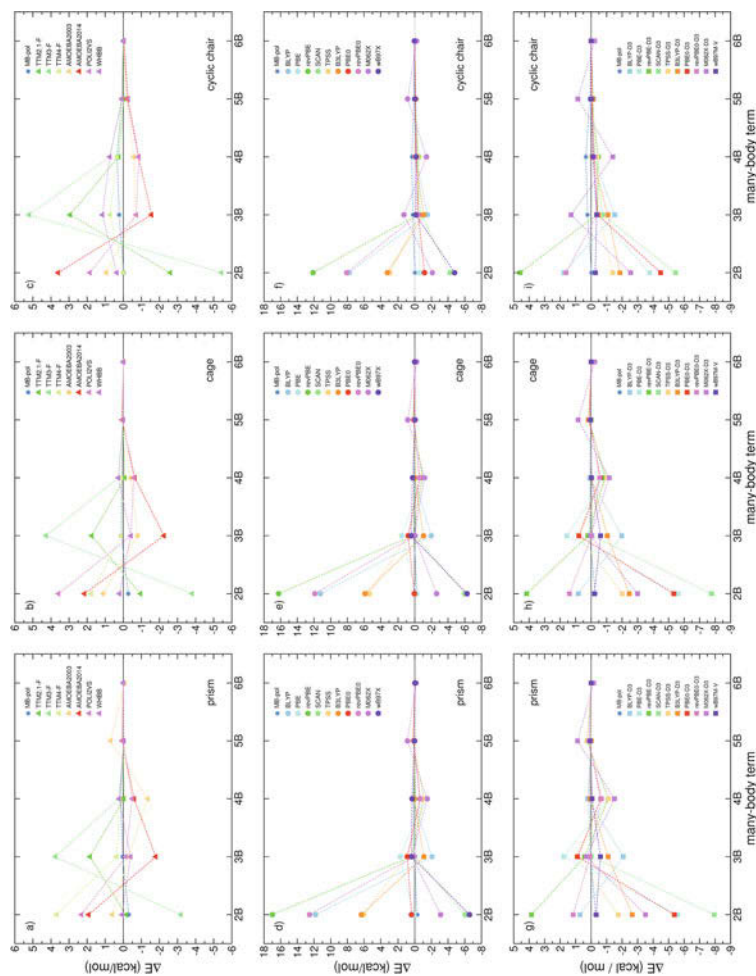
**Fig. 6** Errors, $\left( \Delta E = \Delta E_{nB}^{model} - \Delta E_{nB}^{CCSD(T)} \right)$ relative to CCSD(T)-F12/VTZ-F12 reference values for the individual terms (nB) of the many-body expansion of the interaction energy calculated using MB-pol, polarizable force fields, and DFT models for the (a–c) prism, (d–f) cage, and (g–i) cyclic chair hexamer isomers. The first column (panels a, d, g) reports the results obtained with the polarizable force fields. The second (panels b, e, h) and third (panels c, f, and i) columns report the results obtained with the same DFT models, without and with dispersion corrections, respectively, as in Fig. 5

**Table 1** Space groups and supercell dimensions along with the number of water molecules (N) used in the calculations of the lattice energies (in kcal/mol) for the experimental crystal structures of several ice phases (Pham et al. 2017)

| Ice phase | Space group | Supercell | N | Lattice energy (kcal/mol) | |
|---|---|---|---|---|---|
| | | | | MB-pol | experiment |
| $I_h$ | $P6_3cm$ | (3,3,3) | 324 | −14.09 | −14.07 |
| IX | $P4_12_12$ | (3,3,3) | 324 | −13.86 | −13.97 |
| II | $R\overline{3}$ | (4,4,4) | 768 | −13.90 | −14.04 |
| VIII | $I4_1/amd$ | (4,4,3) | 384 | −13.60 | −13.30 |

calculated using the experimental crystal structures are in good agreement with the corresponding experimental values. Further analysis of the MB-pol results in terms of fundamental energy contributions showed that the differences in lattice energies between different ice phases are primarily dependent on the balance between short-range 2B and 3B interactions, many-body induction, and dispersion energy.

To provide an overall assessment of the MB-pol accuracy, Table 2 reports the results obtained by applying the same scoring scheme that was introduced to compare the ability of various DFT models to predict key properties of the water monomer, dimer, and hexamer and different ice phases (Gillan et al. 2016). Specifically, the properties considered in this analysis are: the harmonic frequency of the monomer symmetric stretch $\left(f_{ss}^{mono}\right)$, the dimer binding energy $\left(E_b^{dim}\right)$, the binding energy per monomer of the cyclic chair hexamer isomer $\left(E_b^{ring}\right)$, the sublimation energy of ice $I_h$ $\left(E_{sub}^{I_h}\right)$, the difference per monomer between the binding energies of the prism and cyclic chair hexamer isomers $\left(\Delta E_b^{prism-ring}\right)$, the difference of sublimation energies of ice $I_h$ and ice VIII $\left(\Delta E_{sub}^{I_h-VIII}\right)$, the equilibrium oxygen-oxygen distance of the dimer $\left(R_{OO}^{dim}\right)$, and the equilibrium volumes per monomer of ice $I_h$ $\left(V_{eq}^{I_h}\right)$ and ice VIII $\left(V_{eq}^{VIII}\right)$. The scores are then assigned by determining the deviations from the corresponding reference data obtained from high-level electronic structure calculations or experimental measurements. A score of 100% is assigned if the magnitude of the deviation is less than a predefined tolerance $\delta x_{tol}$, and a deduction of 10% is applied for each successive increment $\delta x_{tol}$ in $|x - x_{ref}|$. A score of zero is given if $|x - x_{ref}| > 11 \ \delta x_{tol}$. Specific details about the scoring scheme are given in the original study (Gillan et al. 2016). As shown in Table 2, MB-pol scores 90% or higher for all properties except for the difference in binding energies between the prism and cyclic-chair hexamer isomers and the equilibrium volume per monomer of ice VIII. MB-pol's average percentage score is 91% using the reference values reported in the original study (Gillan et al. 2016), which becomes 93% if more accurate reference values for the harmonic frequency of the monomer symmetric stretch and oxygen-oxygen distance in the water dimer are considered. To put things in perspective, the best DFT model among those analyzed in the original study scores 74% (Gillan et al. 2016).

**Table 2** Analysis of the performance of MB-pol and various DFT models in predicting several properties of water according to the scoring scheme introduced by (Gillan et al. 2016). If not indicated otherwise, all reference values and DFT scores are taken from the original study (Gillan et al. 2016). Also listed as reference values (second entries) are the harmonic frequency of the monomer symmetric stretch and oxygen-oxygen distance in the water dimer calculated by Howard et al. (2014). For MB-pol, the first entry corresponds to the value calculated for each property, while the second and, when available, third entries are the percentage scores relative to the corresponding reference values. Specific details about the scoring scheme along with a complete discussion of the DFT results are reported in the original study (Gillan et al. 2016)

| Model | $f_{ss}^{mono}$ (cm$^{-1}$) | $E_b^{dim}$ (meV) | $E_b^{ring}$ (meV) | $E_{sub}^{1h}$ (meV) | $\Delta E_b^{prism-ring}$ (meV) | $\Delta E_b^{1h-VIII}$ (meV) | $R_{OO}^{dim}$ (Å) | $V_{eq}^{1h}$ (Å$^3$) | $V_{eq}^{VIII}$ (Å$^3$) | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Reference | 38123835 | 5.02 | 7.36 | 14.07 | 1.80 | 0.77 | 2.9092.9127 | 32 | 19.1 | |
| MB-pol | 3833 | 4.96 | 7.14 | 14.09 | 2.43 | 0.49 | 2.92 | 31.61 | 18.64 | |
| | 90100 | 100 | 100 | 100 | 80 | 90 | 90100 | 90 | 80 | 91 93 |
| LDA | 60 | 0 | – | 0 | – | 10 | 0 | – | – | 14 |
| PBE | 50 | 100 | 80 | 80 | 0 | 0 | 100 | 70 | 20 | 56 |
| BLYP | 20 | 70 | 80 | 50 | 0 | 0 | 60 | 100 | 0 | 42 |
| PBE0 | 80 | 100 | 90 | 90 | 0 | 0 | 90 | 70 | 40 | 62 |
| revPBE-DRSLL | 30 | 70 | 60 | 50 | 100 | 100 | 0 | 30 | 0 | 49 |
| optPBE-DRSLL | 40 | 100 | 100 | 50 | 100 | 100 | 60 | 90 | 30 | 74 |
| optB88-DRSLL | 60 | 100 | 90 | 20 | 100 | 100 | 50 | 50 | 100 | 74 |
| rPW86-DF2 | 20 | 100 | 100 | 100 | 100 | 100 | 40 | 50 | 0 | 68 |
| PBE-TS | 50 | 80 | 60 | 0 | 100 | 40 | 90 | 30 | 50 | 56 |
| PBE0-TS | 80 | 90 | 80 | 40 | 100 | 60 | 90 | 40 | 70 | 72 |
| BLYP-D3 | 20 | 100 | 90 | 30 | 100 | 40 | 70 | 50 | 90 | 66 |

# 4    From the Born-Oppenheimer Potential Energy Surface to Experimental Observables

Considering the accuracy with which MB-pol predicts the energetics of water from the gas to the condensed phase, there is reason to wonder if this is sufficient for equally accurate predictions of structural, thermodynamic, dynamical, and spectroscopic properties that can be measured experimentally. Since, by construction, MB-pol effectively provides an analytical representation of the Born-Oppenheimer potential energy surface of water, from a rigorous theoretical standpoint, computer simulations with MB-pol should explicitly account for nuclear quantum effects (Tuckerman 2010). For small clusters, methods based on basis set expansions and finite grids have been shown to be quite effective in describing nuclear quantum dynamics with MB-pol (Babin et al. 2013; Richardson et al. 2016; Brown et al. 2017). For systems with many degrees of freedom, methods based on Feynman's path-integral formalism (Tuckerman 2010), such as path-integral molecular dynamics (PIMD), enable numerically exact calculations of structural and thermodynamic properties for a given a Born-Oppenheimer potential energy surface. To date, exact calculations of quantum dynamical properties (e.g., diffusion and vibrational spectra) are unaffordable for systems with many degrees of freedom. Several approximate quantum dynamics methods for condensed phase systems have been proposed, including centroid molecular dynamics (CMD) (Voth 1996), which has been used extensively in quantum simulations of liquid water and ice using MB-pol (Medders et al. 2014, Medders and Paesani 2015, 2016; Straight and Paesani 2016; Moberg et al. 2017; Reddy et al. 2017).

It was shown that MB-pol quantitatively reproduces the experimental vibration-rotation tunneling spectrum of the water dimer as well as both second and third virial coefficients (Babin et al. 2013, 2014). When combined with PIMD simulations carried out at different temperatures within the replica exchange formalism, it was shown that MB-pol correctly reproduces quantum equilibria between different hexamer isomers, predicting cage and prism to be the (quantum) ground-state structures of $(H_2O)_6$ and $(D_2O)_6$, respectively (Brown et al. 2017). Local-monomer/local-mode calculations (Wang and Bowman 2011; Cheng and Steele 2014; Cheng et al. 2016) were then carried out on the isomeric quantum distributions to monitor the microscopic melting of the water hexamer through the analysis of vibrational spectra and various structural order parameters as a function of temperature (Brown et al. 2017). In two joint experimental-theoretical studies, MB-pol was used to determine the tunneling pathways in the prism isomer of the hexamer cluster (Richardson et al. 2016) as well as to interpret the THz spectra of the octamer cluster (Cole et al. 2016).

Classical MD and quantum PIMD and CMD simulations were also carried out to investigate the properties of liquid water (Medders et al. 2014). Classical and quantum oxygen-oxygen (O-O), oxygen-hydrogen (O-H), and hydrogen-hydrogen (H-H) radial distribution functions (RDFs) calculated, respectively, from MD and PIMD simulations in the isobaric-isothermal (constant number of molecules – constant pressure – constant temperature or NPT) ensemble are compared in Fig. 6 with the corresponding experimental data. Nearly quantitative

agreement is obtained between PIMD and experimental O-O RDFs, indicating that MB-pol predicts the correct structure of liquid water and nuclear quantum effects play a nonnegligble role in determining the properties of the underlying hydrogen-bond network. The importance of nuclear quantum effects becomes more apparent from the analysis of the O-H and H-H RDFs, which are predicted to be significantly more structured by classical MD simulations. PIMD simulations at 298.15 K also predict for liquid water a density of 1.001 g cm$^{-3}$ and an enthalpy of vaporization of 10.1 $\pm$ 0.4 kcal mol$^{-1}$, which are in good agreement with the corresponding experimental values of 0.997 g cm$^{-3}$ and 10.5 kcal mol$^{-1}$, respectively. Furthermore, the diffusion coefficient (D) and orientational relaxation time ($\tau_2$) of liquid water calculated from CMD simulations carried out with MB-pol at 298.15 K were found to be 0.22 $\pm$ 0.03 Å$^2$ ps$^{-1}$ and 2.3 $\pm$ 0.3 ps, respectively, which compare well with the corresponding experimental values of D = 0.23 Å$^2$ ps$^{-1}$ and $\tau_2$ = 2.5 ps.

Classical (MD) and quantum (CMD) simulations were also carried out to model both infrared (IR) and Raman spectra of liquid water (Medders and Paesani 2015; Straight and Paesani 2016; Reddy et al. 2017) and ice (Moberg et al. 2017), as well as the vibrational sum-frequency generation (vSFG) spectrum of the air/water interface (Medders and Paesani 2016; Moberg et al. 2018). Figure 7 shows a comparison between classical and quantum IR spectra of liquid water at 298.15 K and ice at 200 K. In these simulations, MB-pol was combined with a many-body representation of the associated dipole moment (MB-$\mu$) surface to calculate the relevant dipole-dipole correlation function (Medders and Paesani 2013, 2015). Both CMD spectra are in good agreement with the experimental data. By contrast, neglecting zero-point energy effects, the classical spectra are blueshifted compared to experiment, particularly in the high frequency region corresponding to the OH stretching band. As expected from the inverse relationship between temperature and De Broglie's thermal wavelength, the magnitude of the blueshift increases from liquid water at 298.15 K to ice at 200 K. A systematic analysis of the IR lineshapes for both liquid water and ice demonstrated that a rigorous representation of many-body effects is necessary for a quantitative reproduction of the experimental data.

Despite the overall agreement between experimental and simulated vibrational spectra, some differences still remain. In particular, the simulated OH stretching band of the IR spectrum of liquid water misses some intensity on the lower frequency portion. It has been argued that this discrepancy is primarily due to the inability of CMD to quantitatively capture the Fermi resonances between the bending overtone and the stretching vibrations (Medders and Paesani 2015; Straight and Paesani 2016). However, further investigations are needed. Second, the positions of the simulated IR OH stretching bands for both liquid water and ice are slightly blueshifted (60 cm$^{-1}$) compared to experiment, indicating that MB-pol predicts slightly too weaker hydrogen bonds. Finally, it was shown that although the vSFG spectrum obtained from CMD simulations with MB-pol correctly reproduces all spectral features in the OH stretching region, the calculated lineshape corresponding to dangling OH bonds is noticeably broader than in the experimental spectra (Medders and Paesani 2016). This deficiency is
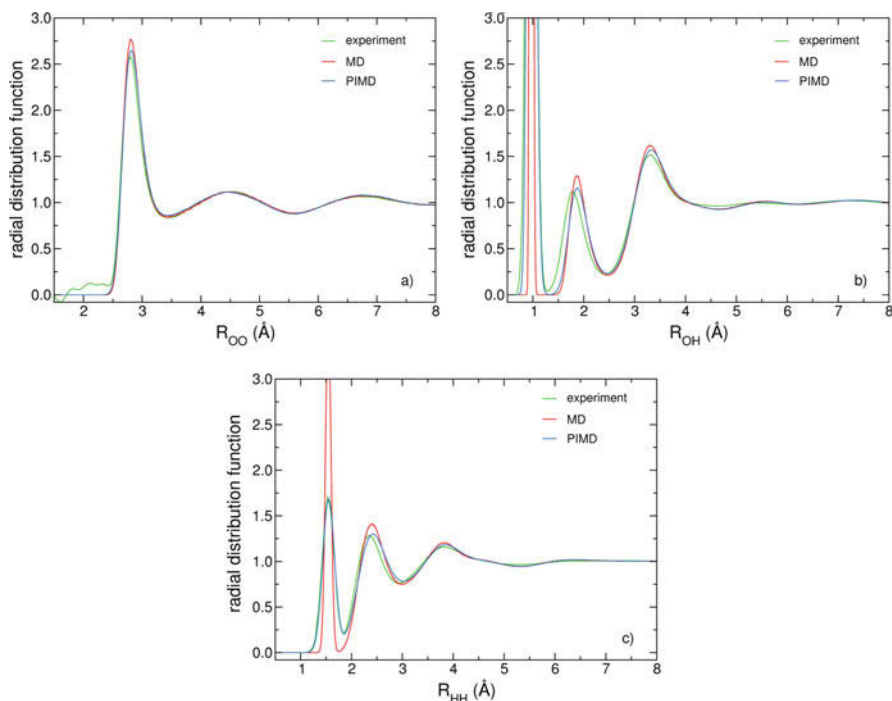
**Fig. 7** Classical (MD) and quantum (PIMD) radial distribution functions calculated for liquid water at 298.15 K. (**a**) O-O, (**b**) O-H, and (**c**) H-H RDFs. The experimental data are from (Skinner et al. 2013) and (Soper and Benmore 2008)

a typical manifestation of approximate quantum dynamics methods based on the path-integral formalism, such as CMD, which, under particular conditions, affects the lineshape of high-frequency vibrational modes (Witt et al. 2009; Paesani and Voth 2010; Rossi et al. 2014) (Fig. 8).

## 5   Summary and Outlook

Building upon the pioneering studies by Stillinger and coworkers (Hankins et al. 1970) and Clementi and coworkers (Matsuoka et al. 1976; Evans et al. 1987; Niesar et al. 1990), the last decade has witnessed renewed interest in the development of accurate potential energy functions rigorously derived from many-body expansions of the interaction energies. These efforts have been supported by continued hardware improvements as well as by the development of efficient algorithms for correlated electronic structure methods which nowadays enable routine calculations of interaction energies for small molecular systems, with chemical accuracy. In parallel with these technological and algorithmic developments, there has been tremendous progress in the representation of multidimensional potential energy
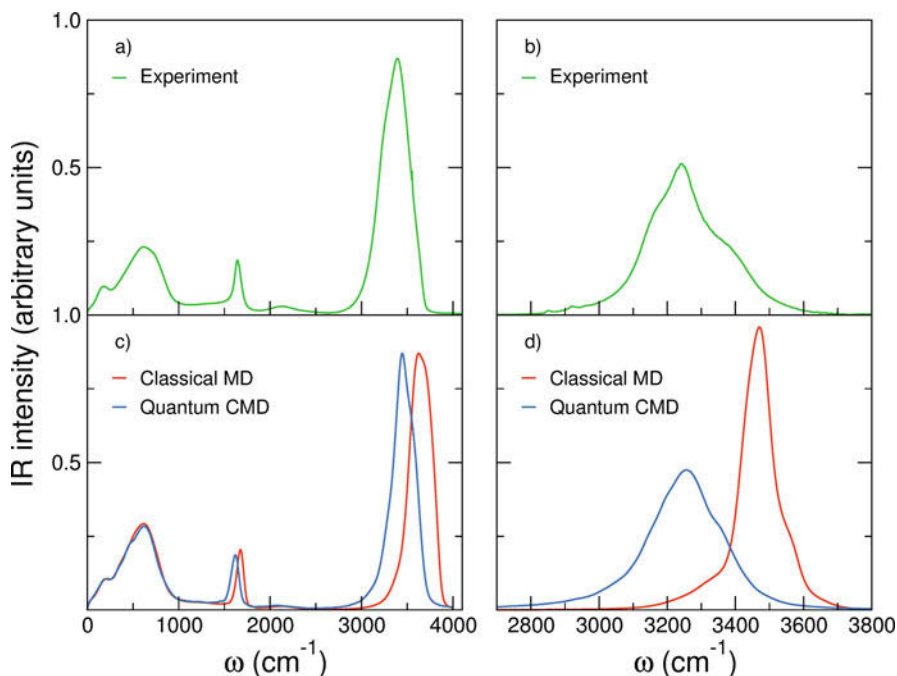
**Fig. 8** Comparisons between experimental (panels a and b) and both classical (MD) and quantum (CMD) simulated (panels c and d) IR spectra of liquid water at 298.15 K (left) and ice at 200 K (right). Adapted from Medders and Paesani (2015), and Moberg et al. (2017)

surfaces with a large number of degrees of freedom, using permutationally invariant polynomials (Braams and Bowman 2009), neural networks (Behler 2016), and Gaussian approximation potentials (Bartók et al. 2010).

Given the importance of water as life's matrix (Ball 2008) and considering the key role that water has played in the history of computer simulations since the first studies by Barker and Watts (1969) and Rahman and Stillinger (Rahman and Stillinger 1971), it is not surprising that several many-body potential energy functions for water have recently been developed, such as CC-pol (Bukowski et al. 2007), WHBB (Wang et al. 2011) HBB2-pol (Babin et al. 2012), and MB-pol (Babin et al. 2013, 2014; Medders et al. 2014). When employed in simulations that allow for explicit treatment of nuclear quantum effects, these PEFs have been shown to correctly reproduce the properties of water from the gas to the condensed phase, thus enhancing the predictive ability of computer simulations.

This chapter has provided a critical review of the performance of MB-pol in representing the Born-Oppenheimer potential energy surface of water from the dimer to small clusters, liquid water, and ice (Paesani 2016). It is shown that MB-pol achieves high accuracy by quantitatively reproducing the individual terms of the many-body expansion of the interaction energy through the combination of

explicit short-range representations of 2B and 3B interactions along with a physically correct description of many-body electrostatic interactions at all distances.

Despite representing a step forward toward the long-sought-after "universal model" capable of correctly describing structural, thermodynamic, dynamical, and spectroscopic properties of water across different phases, it should always be kept in mind that MB-pol is still a "model" of water, with its own approximations and limitations. In particular, MB-pol is a physical model of water in the limit $pK_w \rightarrow \infty$, i.e., autoionization events are not allowed. While this is a reasonable approximation of pure water, the situation drastically changes in complex aqueous solutions with pH values lower or higher than pure water. In addition, as demonstrated by the analysis of vibrational spectra, some experimental features are not reproduced quantitatively by quantum dynamics simulations with MB-pol. Although it has been argued that these differences may be related to the approximate nature of quantum dynamics methods, further studies are necessary to address these deficiencies. Finally, the application of MB-pol in studies of complex solutions will require the development of many-body representations for generic molecular systems. Although some progress has already been made along this direction (Bajaj et al. 2016; Riera et al. 2017; Sode and Cherry 2017; Wang and Bowman 2017), further theoretical and computational developments are needed. In this context, given the unconventional functional form adopted by MB-pol, which is not supported by popular software for computer simulations, synergistic efforts between theoretical/computational chemists/physicists and computer scientists will be key to the development of specialized and more efficient software for many-body molecular dynamics simulations which can take full advantage of modern hardware.

# References

Abu-Mostafa YS, Magdon-Ismail M, Lin H-T (2012) Learning from Data. AML Book

Adamo C, Barone V (1999) Toward reliable density functional methods without adjustable parameters: the PBE0 model. J Chem Phys 110:6158–6170

Adler TB, Knizia G, Werner HJ (2007) A simple and efficient CCSD(T)-F12 approximation. J Chem Phys 127:221106

Babin V, Medders GR, Paesani F (2012) Toward a universal water model: first principles simulations from the dimer to the liquid phase. J Phys Chem Lett 3:3765–3769

Babin V, Leforestier C, Paesani F (2013) Development of a "first principles" water potential with flexible monomers: dimer potential energy surface, VRT spectrum, and second virial coefficient. J Chem Theory Comput 9:5395–5403

Babin V, Medders GR, Paesani F (2014) Development of a "first principles" water potential with flexible monomers. II: trimer potential energy surface, third virial coefficient, and small clusters. J Chem Theory Comput 10:1599–1607

Bajaj P, Götz AW, Paesani F (2016) Toward chemical accuracy in the description of ion-water interactions through many-body representations. I. Halide-water dimer potential energy surfaces. J Chem Theory Comput 12:2698–2705

Ball P (2008) Water as an active constituent in cell biology. Chem Rev 108:74–108

Barker JA, Watts RO (1969) Structure of water; a Monte Carlo calculation. Chem Phys Lett 3: 144–145

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403

Bates DM, Tschumper GS (2009) CCSD(T) complete basis set limit relative energies for low-lying water hexamer structures. J Phys Chem A 113:3555–3559

Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38:3098–3100

Becke AD (1993) Density-functional thermochemistry. 3. The role of exact exchange. J Chem Phys 98:5648–5652

Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145:170901

Braams BJ, Bowman JM (2009) Permutationally invariant potential energy surfaces in high dimensionality. Int Rev Phys Chem 28:577–606

Brown SE, Götz AW, Cheng X, Steele RP, Mandelshtam VA et al (2017) Monitoring water clusters "melt" through vibrational spectroscopy. J Am Chem Soc 139:7082–7088

Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A (2007) Predictions of the properties of water from first principles. Science 315:1249–1252

Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A (2008) Polarizable interaction potential for water from coupled cluster calculations. II. Applications to dimer spectra, virial coefficients, and simulations of liquid water. J Chem Phys 128:094314

Burnham CJ, Anick DJ, Mankoo PK, Reiter GF (2008) The vibrational proton potential in bulk liquid water and ice. J Chem Phys 128:154519

Chai JD, Head-Gordon M (2008) Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. Phys Chem Chem Phys 10:6615–6620

Cheng XL, Steele RP (2014) Efficient anharmonic vibrational spectroscopy for large molecules using local-mode coordinates. J Chem Phys 141:104105

Cheng XL, Talbot JJ, Steele RP (2016) Tuning vibrational mode localization with frequency windowing. J Chem Phys 145:124112

Cisneros GA, Piquemal J-P, Darden TA (2006) Quantum mechanics/molecular mechanics electro-static embedding with continuous and discrete functions. J Phys Chem B 110:13682–13684

Cisneros GA, Elking D, Piquemal J-P, Darden TA (2007) Numerical fitting of molecular properties to Hermite gaussians. J Phys Chem A 111:12049–12056

Cisneros GA, Wikfeldt KT, Ojamae L, Lu JB, Xu Y et al (2016) Modeling molecular interactions in water: from pairwise to many-body potential energy functions. Chem Rev 116:7501–7528

Cole WTS, Farrell JD, Wales DJ, Saykally RJ (2016) Structure and torsional dynamics of the water octamer from THz laser spectroscopy near 215 mu m. Science 352:1194–1197

Dunning TH (1989) Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. J Chem Phys 90:1007–1023

Eisenberg D, Kauzmann W (1969) The structure and properties of water. Oxford University Press, Oxford

Evans MW, Refson K, Swamy KN, Lie GC, Clementi E (1987) Molecular dynamics simulation of liquid water with an ab initio flexible water-water interaction potential. 2. The effect of internal vibrations on the time correlation functions. Phys Rev A 36:3935–3942

Fanourgakis GS, Xantheas SS (2008) Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable, Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. J Chem Phys 128:074506

Frank HS, Wen WY (1957) Structural aspects of ion-solvent interaction in aqueous solutions – a suggested picture of water structure. Discuss Faraday Soc 24:133–140

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA et al (2009) Gaussian 09. Gaussian, Inc., Wallingford

Gao JL, Truhlar DG, Wang YJ, Mazack MJM, Loffler P et al (2014) Explicit polarization: a quantum mechanical framework for developing next generation force fields. Acc Chem Res 47:2837–2845

Gillan MJ, Alfè D, Michaelides A (2016) Perspective: how good is DFT for water? J Chem Phys 144:130901

Goerigk L, Grimme S (2011) A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. Phys Chem Chem Phys 13:6670–6688

Gora U, Podeszwa R, Cencek W, Szalewicz K (2011) Interaction energies of large clusters from many-body expansion. J Chem Phys 135:224102

Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys 132:154104

Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. J Mol Liq 101:219–260

Halkier A, Helgaker T, Jorgensen P, Klopper W, Olsen J (1999a) Basis-set convergence of the energy in molecular Hartree-Fock calculations. Chem Phys Lett 302:437–446

Halkier A, Klopper W, Helgaker T, Jorgensen P, Taylor PR (1999b) Basis set convergence of the interaction energy of hydrogen-bonded complexes. J Chem Phys 111:9157–9167

Handley CM, Hawe GI, Kell DB, Popelier PLA (2009) Optimal construction of a fast and accurate polarisable water potential based on multipole moments trained by machine learning. Phys Chem Chem Phys 11:6365–6376

Hankins D, Moskowitz JW, Stillinger FH (1970) Water molecule interactions. J Chem Phys 53:4544

Hasegawa T, Tanimura Y (2011) A polarizable water model for intramolecular and intermolecular vibrational spectroscopies. J Phys Chem B 115:5545–5553

Howard JC, Gray JL, Hardwick AJ, Nguyen LT, Tschumper GS (2014) Getting down to the fundamentals of hydrogen bonding: anharmonic vibrational frequencies of $(HF)_2$ and $(H_2O)_2$ from ab initio electronic structure computations. J Chem Theory Comput 10:5426–5435

Khaliullin RZ, Cobar EA, Lochan RC, Bell AT, Head-Gordon M (2007) Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. J Phys Chem A 111:8753–8765

Khaliullin RZ, Bell AT, Head-Gordon M (2008) Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals. J Chem Phys 128:184112

Knizia G, Adler TB, Werner HJ (2009) Simplified CCSD(T)-F12 methods: theory and benchmarks. J Chem Phys 130:054104

Manna D, Kesharwani MK, Sylvetsky N, Martin JML (2017) Conventional and explicitly correlated ab initio benchmark study on water clusters: revision of the BEGDB and WATER27 data sets. J Chem Theory Comput 13:3136–3152

Mardirossian N, Head-Gordon M (2015) Mapping the genome of meta-generalized gradient approximation density functionals: the search for B97M-V. J Chem Phys 142:074111

Maréchal Y (2007) The hydrogen bond and the water molecule: the physics and chemistry of water, aqueous and bio media. Elsevier, Amsterdam

Matsuoka O, Clementi E, Yoshimine M (1976) CI study of the water dimer potential surface. J Chem Phys 64:1351–1361

Medders GR, Paesani F (2013) Many-body convergence of the electrostatic properties of water. J Chem Theory Comput 9:4844–4852

Medders GR, Paesani F (2015) Infrared and Raman spectroscopy of liquid water through "first-principles" many-body molecular dynamics. J Chem Theory Comput 11:1145–1154

Medders GR, Paesani F (2016) Dissecting the molecular structure of the air/water interface from quantum simulations of the sum-frequency generation spectrum. J Am Chem Soc 138:3912–3919

Medders GR, Babin V, Paesani F (2014) Development of a "first-principles" water potential with flexible monomers. III. Liquid phase properties. J Chem Theory Comput 10:2906–2910

Medders GR, Götz AW, Morales MA, Bajaj P, Paesani F (2015) On the representation of many-body interactions in water. J Chem Phys 143:104102

Moberg DR, Straight SC, Knight C, Paesani F (2017) Molecular origin of the vibrational structure of ice I$_h$. J Phys Chem Lett 8:2579–2583

Moberg DR, Straight SC, Paesani F et al (2018) Temperature dependence of the air/water interface revealed by polarization sensitive sum-frequency generation spectroscopy. J. Phys. Chem. B 122:4356–4365

Morales MA, Gergely JR, McMinis J, McMahon JM, Kim J et al (2014) Quantum Monte Carlo benchmark of exchange-correlation functionals for bulk water. J Chem Theory Comput 10:2355–2362

Niesar U, Corongiu G, Clementi E, Kneller GR, Bhattacharya DK (1990) Molecular dynamics simulations of liquid water using the NCC ab initio potential. J Phys Chem 94:7949–7956

Paesani F (2016) Getting the right answers for the right reasons: toward predictive molecular simulations of water with many-body potential energy functions. Acc Chem Res 49: 1844–1851

Paesani F, Voth GA (2010) A quantitative assessment of the accuracy of centroid molecular dynamics for the calculation of the infrared spectrum of liquid water. J Chem Phys 132: 014105

Partridge H, Schwenke DW (1997) The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. J Chem Phys 106:4618–4639

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Peterson KA, Adler TB, Werner HJ (2008) Systematically convergent basis sets for explicitly correlated wavefunctions: the atoms H, He, B-Ne, and Al-Ar. J Chem Phys 128:084102

Pham CH, Reddy SK, Chen K, Knight C, Paesani F (2017) Many-body interactions in ice. J Chem Theory Comput 13:1778–1784

Piquemal JP, Cisneros GA, Reinhardt P, Gresh N, Darden TA (2006) Towards a force field based on density fitting. J Chem Phys 124:104101

Rahman A, Stillinger FH (1971) Molecular dynamics study of liquid water. J Chem Phys 55:3336

Rappoport D, Furche F (2010) Property-optimized Gaussian basis sets for molecular response calculations. J Chem Phys 133:134105

Reddy SK, Straight SC, Bajaj P, Pham CH, Riera M et al (2016) On the accuracy of the MB-pol many-body potential for water: interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. J Chem Phys 145:194504

Reddy SK, Moberg DR, Straight SC, Paesani F (2017) Temperature-dependent vibrational spectra and structure of liquid water from classical and quantum simulations with the MB-pol potential energy function. J Chem Phys 147:244504

Ren PY, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107:5933–5947

Rezac J, Hobza P (2013) Describing noncovalent interactions beyond the common approximations: how accurate is the "gold standard," CCSD(T) at the complete basis set limit? J Chem Theory Comput 9:2151–2155

Richardson JO, Perez C, Lobsiger S, Reid AA, Temelso B et al (2016) Concerted hydrogen-bond breaking by quantum tunneling in the water hexamer prism. Science 351:1310–1313

Riera M, Mardirossian N, Bajaj P, Götz AW, Paesani F (2017) Toward chemical accuracy in the description of ion-water interactions through many-body representations. Alkali-water dimer potential energy surfaces. J Chem Phys 147:161715

Rossi M, Liu HC, Paesani F, Bowman J, Ceriotti M (2014) Communication: on the consistency of approximate quantum dynamics simulation methods for vibrational spectra in the condensed phase. J Chem Phys 141:181101

Shao YH, Gan ZT, Epifanovsky E, Gilbert ATB, Wormit M et al (2015) Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. Mol Phys 113:184–215

Shvab I, Sadus RJ (2016) Atomistic water models: aqueous thermodynamic properties from ambient to supercritical conditions. Fluid Phase Equilib 407:7–30

Simova L, Rezac J, Hobza P (2013) Convergence of the interaction energies in noncovalent complexes in the coupled-cluster methods up to full configuration interaction. J Chem Theory Comput 9:3420–3428

Skinner LB, Huang CC, Schlesinger D, Pettersson LGM, Nilsson A et al (2013) Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. J Chem Phys 138:074506

Sode O, Cherry JN (2017) Development of a flexible-monomer two-body carbon dioxide potential and its application to clusters up to $(CO_2)_{13}$. J Comput Chem 38:2763–2774

Soper AK, Benmore CJ (2008) Quantum differences between heavy and light water. Phys Rev Lett 101:065502

Stone AJ (1997) The theory of intermolecular forces. Clarendon Press, Oxford

Straight SC, Paesani F (2016) Exploring electrostatic effects on the hydrogen bond network of liquid water through many-body molecular dynamics. J Phys Chem B 120:8539–8546

Sun JW, Ruzsinszky A, Perdew JP (2015) Strongly constrained and appropriately normed semilocal density functional. Phys Rev Lett 115:036402

Tao FM, Pan YK (1992) Moller-Plesset perturbation investigation of the $He_2$ potential and the role of midbond basis functions. J Chem Phys 97:4989–4995

Tao JM, Perdew JP, Staroverov VN, Scuseria GE (2003) Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. Phys Rev Lett 91:146401

Temelso B, Archer KA, Shields GC (2011) Benchmark structures and binding energies of small water clusters with anharmonicity corrections. J Phys Chem A 115:12034–12046

Thole BT (1981) Molecular polarizabilities calculated with a modified dipole interaction. Chem Phys 59:341–350

Torheyden M, Jansen G (2006) A new potential energy surface for the water dimer obtained from separate fits of ab initio electrostatic, induction, dispersion and exchange energy contributions. Mol Phys 104:2101–2138

Tuckerman ME (2010) Statistical mechanics: theory and molecular simulation. Oxford University Press, Oxford

van Duijnen PT, Swart M (1998) Molecular and atomic polarizabilities: Thole's model revisited. J Phys Chem A 102:2399–2407

Vega C, Abascal JLF (2011) Simulating water with rigid non-polarizable models: a general perspective. Phys Chem Chem Phys 13:19663–19688

Voth GA (1996) Path-integral centroid methods in quantum statistical mechanics and dynamics. Adv Chem Phys 93:135–218

Vydrov OA, Van Voorhis T (2010) Nonlocal van der Waals density functional: the simpler the better. J Chem Phys 133:244103

Wang LP, Head-Gordon T, Ponder JW, Ren P, Chodera JD et al (2013) Systematic improvement of a classical molecular model of water. J Phys Chem B 117:9956–9972

Wang QK, Bowman JM (2017) Two-component, ab initio potential energy surface for $CO_2$-$H_2O$, extension to the hydrate clathrate, $CO_2@(H_2O)_{20}$, and VSCF/VCI vibrational analyses of both. J Chem Phys 147:161714

Wang YM, Bowman JM (2011) Ab initio potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters. J Chem Phys 134:154510

Wang YM, Huang XC, Shepler BC, Braams BJ, Bowman JM (2011) Flexible, ab initio potential, and dipole moment surfaces for water. I. Tests and applications for clusters up to the 22-mer. J Chem Phys 134:094509

Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. Phys Chem Chem Phys 7:3297–3305

Witt A, Ivanov SD, Shiga M, Forbert H, Marx D (2009) On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. J Chem Phys 130:194510

Zhang YK, Yang WT (1998) Comment on "generalized gradient approximation made simple". Phys Rev Lett 80:890

Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor Chem Accounts 120:215–241

# Neural Network Potentials in Materials Modeling

# 31

## Matti Hellström and Jörg Behler

## Contents

**Abstract**

The availability of reliable interatomic potentials is necessary for carrying out computer simulations of complex materials. While electronic structure methods like density functional theory have been applied with great success to many systems, the high computational costs of these methods severely restrict the scientific problems that can be studied. Consequently, in recent years a lot of effort has been spent on the development of more efficient potentials enabling large-scale simulations. In particular, machine learning potentials have received considerable attention, because they promise to combine the accuracy of first-principles methods with the efficiency of force fields. In this chapter an important

M. Hellström (✉) · J. Behler (✉)
Institut für Physikalische Chemie, Theoretische Chemie, Universität Göttingen, Göttingen, Germany
e-mail: matti.hellstroem@chemie.uni-goettingen.de; joerg.behler@uni-goettingen.de

class of machine learning potentials employing artificial neural networks will be reviewed and discussed.

# 1 Introduction

Artificial neural networks (NNs) are phenomenally versatile and are at present used in a wide range of machine learning (ML) applications (Bishop 1996; Haykin 2011): speech and hand-writing recognition, self-driving vehicles, business intelligence, industrial process control, and game playing, to name just a few. Neural networks are capable of handling the two major types of problems targeted by ML algorithms: classification and regression. This chapter focuses on how NNs can be used for a particular regression problem, namely, the prediction of the potential energy surface (PES), i.e., the potential energy as a function of the atomic positions in a system. Such a neural network is referred to as a neural network potential (NNPs) (Handley and Popelier 2010; Behler 2011b, 2017).

The PES lies at the heart of many problems in materials modeling, theoretical condensed matter physics, and computational chemistry. It gives the relative stabilities of different atomic configurations and is directly connected to, for example, mechanical properties, defect distributions, reaction rates, thermodynamic equilibria, spectroscopic signatures, and many other properties.

There are multiple ways of calculating the PES, where the most commonly used in the materials modeling community include electronic structure methods, most prominently density functional theory (DFT) (Parr and Yang 1989), and a wide range of atomistic or even coarse-grained potentials. DFT calculations have been proven to have a fairly good predictive capability for many experimental properties. However, DFT calculations become computationally very demanding if large systems are modeled. This typically limits the applicability of DFT to static (single-point) calculations of a few hundred atoms or to short dynamic simulations in the order of several hundred picoseconds. Atomistic potentials, on the other hand, are more approximate and thus provide estimations for the PES at a much lower computational cost than DFT. They can therefore practically be applied to large-scale simulations. The required parameters are typically fitted to reproduce either some experimental properties of the system modeled or to reproduce some key results like energies from DFT calculations. The latter approach is an example of the so-called multiscale modeling, where information from a high level of theory (DFT) is used to parameterize a lower level of theory.

A neural network potential is in many ways similar to a force field. However, unlike force fields, the functional forms of NNPs are not based on any physical approximations. Instead, the great flexibility of NNs is exploited, and the NNP is parameterized to reproduce the PES obtained from DFT reference calculations or from some other electronic structure method. The NNP then provides a computationally inexpensive way to predict the PES, which makes it possible to use NNPs in large-scale Monte Carlo and molecular dynamics simulations to sample configurational and phase space. During the NNP parameterization, the NN "learns"

the stabilities of different molecules and structural motifs. An NNP is thus inherently "reactive," meaning that chemical reactions or significant structural rearrangements including the breaking and forming of covalent bonds can be accurately described, as long as the relevant training data is provided. To the end user, there is no apparent difference between using a neural network potential or a reactive force field. Both types of approaches provide the PES at a similar computational cost and can be applied for similar types of reactive systems.

NNs have been used for more than two decades for the representation of potential energy surfaces starting with the pioneering work of Doren and coworkers (Blank et al. 1995). The first generation of NNPs suffered from a restriction to small molecular systems containing only a few atoms, but this limitation could be overcome in 2007 by the introduction of high-dimensional NNPs (Behler and Parrinello 2007). In recent years also ML methods other than neural networks have found use in a similar fashion to construct potentials for atomistic simulations of complex systems. For example, methods like Gaussian approximation potentials (Bartók et al. 2010), kernel ridge regression (Rupp et al. 2012), and support vector machines (Balabin and Lomakina 2011) can also be used to describe the PES and related quantities. A review of recent advances in machine learning-based interatomic potentials was given by Behler (2016).

A presentation of the entire theory of artificial neural networks, which can be found in many text books (e.g., Bishop 1996; Haykin 2011), is beyond the scope of this chapter. We will therefore focus on the design choices that are typically made for estimating the PES of a given material. The potential energy is a single real-valued number, and the input features representing the atomic structure are real-valued numbers, so the NNP is a function $\chi : \mathbb{R}^n \to \mathbb{R}$.

We will discuss how NNPs can be used to compute the potential energy, how the conversion from a set of atomic positions to a suitable NN input can be made, how a NNP is parameterized, and how a NNP is validated, as well as some of the strengths and weaknesses of NNPs.

## 2 A Feed-Forward Neural Network

Figure 1 shows a schematic representation of a small fully connected feed-forward NN that defines a function $\chi : \mathbb{R}^2 \to \mathbb{R}$, which transforms an input vector $\mathbf{G} = (G_1, G_2)^{\mathrm{T}}$ into the output value $E$. The function has several parameters contained in the weight matrices $\mathbf{a}^{(0)}$, $\mathbf{b}^{(0)}$, $\mathbf{a}^{(1)}$, $\mathbf{b}^{(1)}$, $\mathbf{a}^{(2)}$, and $\mathbf{b}^{(2)}$, defining $E = \chi(\mathbf{G}; \mathbf{a}^{(0)}, \mathbf{b}^{(0)}, \mathbf{a}^{(1)}, \mathbf{b}^{(1)}, \mathbf{a}^{(2)}, \mathbf{b}^{(2)})$. The example NN consists of an input layer with two nodes, two hidden layers that contain three nodes each, and an output layer containing one node. In addition, a bias node with a fixed value of 1 is connected to all nodes in the hidden layers and the output layer (shown in blue). The example NN has an architecture of 2–3–3–1. The input features $(G_1, G_2)$ in layer 0 correspond to the values $(y_1^{(0)}, y_2^{(0)})$.
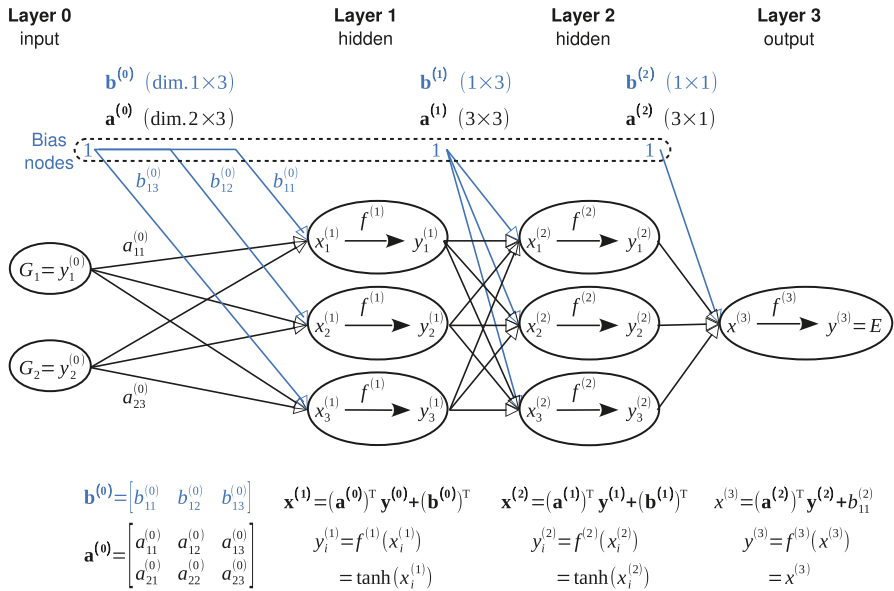
**Fig. 1** Illustration of a 2–3–3–1 feed-forward neural network along with typical choices of the activation functions $f^{(k)}(x)$. This NN defines a function $E = \chi(G_1, G_2; \mathbf{a}^{(0)}, \mathbf{b}^{(0)}, \mathbf{a}^{(1)}, \mathbf{b}^{(1)}, \mathbf{a}^{(2)}, \mathbf{b}^{(2)})$. Black arrows with empty heads between $y_i^{(k)}$ and $x_j^{(k+1)}$ represent the element $a_{ij}^{(k)}$ in the weight matrix $\mathbf{a}^{(k)}$; this is explicitly shown for a few of the elements of $\mathbf{a}^{(0)}$. Similarly, the blue arrows represent the bias weights. Arrows with filled heads represent the application of the activation function $f^{(k)}$

The computation proceeds left to right: from the input layer values, the first hidden layer values $\mathbf{y}^{(1)} = (y_1^{(1)}, y_2^{(1)}, y_3^{(1)})^{\mathrm{T}}$ are computed, and a bias node is added; those values are then used to compute $\mathbf{y}^{(2)} = (y_1^{(2)}, y_2^{(2)}, y_3^{(2)})^{\mathrm{T}}$ in the second hidden layer and a bias node is added; those values are used to compute the final output value $E$.

The values in a preceding layer, $\mathbf{y}^{(k-1)}$, are used to compute the values of $\mathbf{y}^{(k)}$ in the next layer via an intermediate vector $\mathbf{x}^{(k)}$:

$$\mathbf{x}^{(k)} = (\mathbf{a}^{(k-1)})^{\mathrm{T}}\mathbf{y}^{(k-1)} + (\mathbf{b}^{(k-1)})^{\mathrm{T}}. \tag{1}$$

In Eq. 1, each element of the vector $\mathbf{x}^{(k)}$ is calculated as the sum of a bias weight in $(\mathbf{b}^{(k-1)})^{\mathrm{T}}$ and a linear combination of the elements of $\mathbf{y}^{(k-1)}$, with the coefficients given in the weight matrix $\mathbf{a}^{(k-1)}$. For example, in Fig. 1, $x_1^{(1)} = a_{11}^{(0)}y_1^{(0)} + a_{21}^{(0)}y_2^{(0)} + b_{11}^{(0)}$. The elements of $\mathbf{a}^{(k-1)}$ and $\mathbf{b}^{(k-1)}$ are parameters (NN weights) that must be determined before the NN can be used for energy predictions; how this determination is made is shown later. The purpose of the bias node (with the value of 1, shown in blue in Fig. 1) in layer $k-1$ is to add some desired constants that do

not depend on the values $\mathbf{y}^{(k-1)}$, to each element of $\mathbf{x}^{(k)}$ (i.e., to the next layer). The values of the vector $\mathbf{x}^{(k)}$ are then transformed via an *activation function* (sometimes called transfer function or basis function) $f^{(k)}(x)$:

$$y_i^{(k)} = f^{(k)}(x_i^{(k)}) \tag{2}$$

For the hidden layers, a nonlinear sigmoid activation function is typically used, for example, the logistic function $f(x) = \frac{1}{1+\exp(-x)}$ or the hyperbolic tangent $f(x) = \tanh(x)$. There are also other possible choices for the hidden layer activation functions. For the output layer, the linear activation function $f(x) = x$ is used to avoid a constrained range of output values of the NN.

For a NN to predict the potential energy of the system, the input vector $\mathbf{G}$ must contain all of the information about the system which determines the potential energy. In the absence of any external fields, the potential energy is invariant under rotation and translation. Thus, a poor choice would be to simply select the Cartesian coordinates of the atoms as the input vector, since these do not possess these invariances. A better choice would be internal coordinates like interatomic distances and bond angles that by design are invariant under rotation and translation. This works well as long as the number of internal coordinates is manageable, for example, to describe the PES of small molecules. However, for large systems containing thousands or more degrees of freedom, this approach quickly becomes intractable, since the number of input features becomes very large. Another severe challenge is the inclusion of permutation invariance of the PES with respect to the interchange of atoms of the same element, which is also problematic in case of internal coordinates.

Rather than having the NN describe the PES of the entire system (a "low-dimensional" NNP), it is possible to construct NNs that describe the PES of individual atoms. By combining several such atomic NNs together, the number of input features can be limited while retaining the ability to model large systems with thousands of atoms. These high-dimensional NNPs will be the focus of the rest of this chapter. High-dimensional NNPs are typically used in conjunction with the so-called symmetry functions as input features. They are described in the next section.

## 3　High-Dimensional Neural Network Potentials and Symmetry Functions

In a high-dimensional neural network potential, the total energy is computed as the sum of "atomic" energies $E_i$:

$$E = \sum_{i=1}^{N^{\text{atom}}} E_i \tag{3}$$

where $N^{\text{atom}}$ is the number of atoms in the system. Each atomic energy $E_i$ is determined by means of an element-dependent feed-forward NN, where the input features describe the chemical environment around the atom. This is akin to how the sum over atomic environments (e.g., bond lengths and angles) is used to calculate the total energy in many force fields. However, high-dimensional NNs have a considerably more flexible functional form and can easily capture high-order many-body effects; indeed, the input features are typically computed using particular many-body functions known as symmetry functions.

For a multicomponent system with more than one element, one NN is employed for each element. For example, to describe a CuAgAu alloy, three elemental NNs are created, one for Cu, one for Ag, and one for Au, and each NN is called once for each atom of the respective element. Different NN architectures (number of input features, number of hidden layers, and/or number of nodes per hidden layer) can be used for the different elements. The atomic energies for all the Cu atoms in the system would be evaluated using the same Cu-specific NN, with only the numerical values of the input vectors differing, since the atomic environments around different atoms may vary. This treatment ensures that the total energy is permutationally invariant with respect to the order with which the Cu atoms are provided in the input file. Figure 2 shows schematically how the total energy $E$ is calculated using a high-dimensional neural network potential.

Below, we refer to specific atoms using the indices $i$, $j$, and $k$, and to elements using capital letters $I$, $J$, and $K$, and we use the notation $i \in I$ to specify that atom $i$ is of element $I$. The set of all atomic positions together with the corresponding elements is denoted $\{\mathbf{R}, Z\}$, where $\{\mathbf{R}\}$ denotes the Cartesian coordinates and $Z$ the element. Thus, $\{\mathbf{R}, Z\}$ contains all of the information normally contained in a single frame of a Monte Carlo simulation or of a single frame of a molecular dynamics simulation (without the particle velocities).

For an atom $i \in I$, the atomic energy $E_i$ in Eq. 3 is calculated as

$$E_{i \in I} = \chi^I(\mathbf{G}^I(i)) \tag{4}$$

where $\chi^I$ denotes the NN for element $I$. The input vector $\mathbf{G}^I(i)$ is a vector of symmetry function values:

$$\mathbf{G}^I(i \in I) = \begin{pmatrix} G_1^I(i, \{\mathbf{R}, Z\}) \\ G_2^I(i, \{\mathbf{R}, Z\}) \\ \vdots \\ G_{N^{\text{sym}}(I)}^I(i, \{\mathbf{R}, Z\}) \end{pmatrix} \tag{5}$$

The symmetry functions are descriptors of the chemical environment around the atoms. Typically, only the local environment within a cutoff sphere of radius $R_{\text{cut}}$ is considered. This is achieved by means of a tapering function $f_{\text{cut}}(R)$ that smoothly decays to 0 in value and slope at $R = R_{\text{cut}}$. A common choice for $f_{\text{cut}}(R)$ is
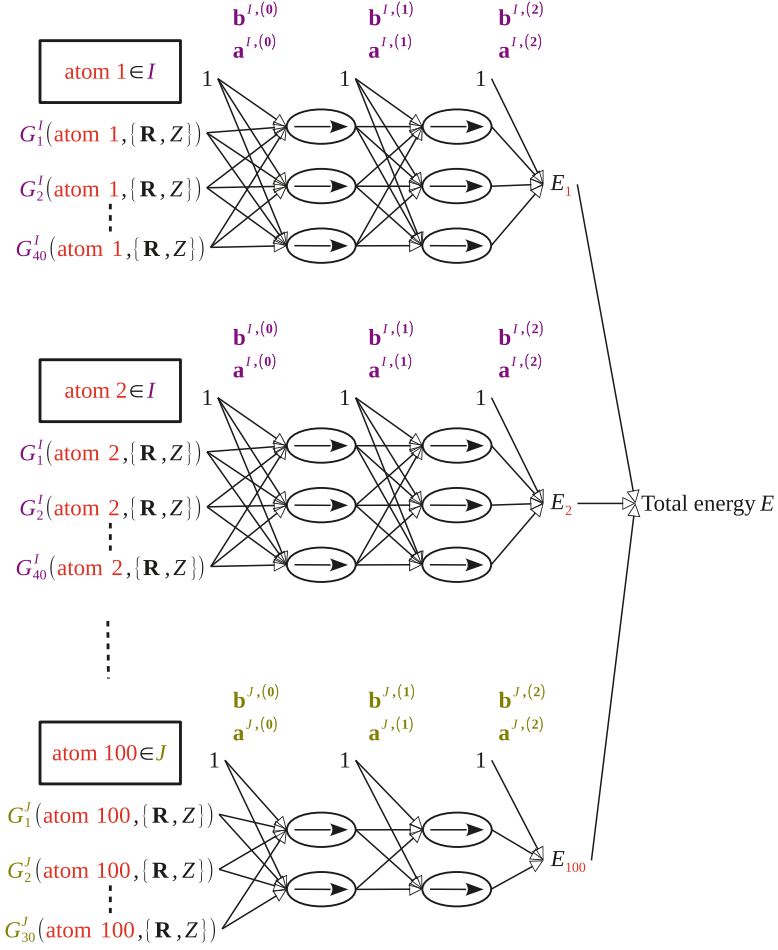
**Fig. 2** Schematic representation of the evaluation of the total energy $E$ for a system with 100 atoms and two elements $I$ and $J$, using a high-dimensional neural network potential. In this example, $N^{\text{sym}}(I) = 40$ and $N^{\text{sym}}(J) = 30$. The evaluation of each atomic energy $E_i$ is performed as in Fig. 1

$$f_{\text{cut}}(R) = \begin{cases} \tanh^3\left(1 - \dfrac{R}{R_{\text{cut}}}\right) & R \leq R_{\text{cut}} \\ 0 & R > R_{\text{cut}} \end{cases} \qquad (6)$$

although other forms of $f_{\text{cut}}$ are also possible. An example of a "radial symmetry" function for an atom $i \in I$ is (Behler 2011a)
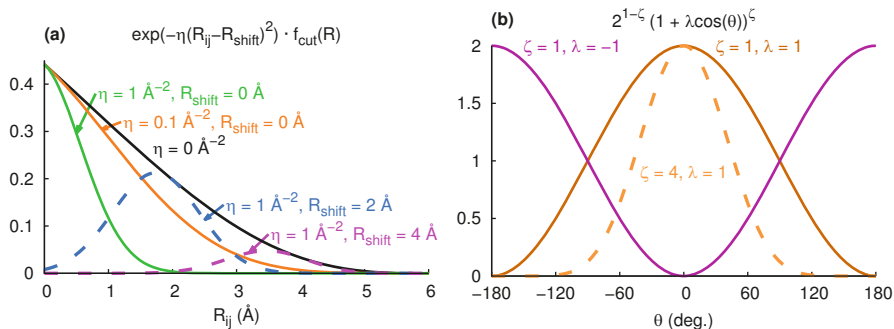
**Fig. 3** (**a**) The summand of the radial symmetry function in Eq. 7 for some different values of $\eta$ and $R_{\text{shift}}$, using the cutoff function $f_{\text{cut}}(R)$ from Eq. 6. (**b**) The angular part of the summand of the symmetry function in Eq. 8, for some selected values of $\zeta$ and $\lambda$

$$G^I(i \in I, \{\mathbf{R}, Z\}; \ \eta, R_{\text{shift}}, J) = \varphi \left( \sum_{\substack{j \in J \\ j \neq i}} e^{-\eta (R_{ij} - R_{\text{shift}})^2} \cdot f_{\text{cut}}(R_{ij}) \right). \tag{7}$$

This function is a descriptor for how atoms of the element $J$ (that can be same as or different from $I$) are distributed around the atom $i \in I$. It is a sum of Gaussians multiplied by the cutoff function $f_{\text{cut}}(R)$. The optional feature scaling function $\varphi$ modifies the range of values output by the symmetry function $G^I$, which can be useful during the NN parameterization. The scaling function is further discussed in Sect. 4.2. The two parameters $\eta$ and $R_{\text{shift}}$ determine the width and center of the Gaussian functions. Figure 3a shows the value of the summand in Eq. 7 for different distances, for a few selected values of $\eta$ and $R_{\text{shift}}$. Here, the cutoff distance is set to the typical value of $R_{\text{cut}} = 6 \,\text{Å}$, and the cutoff function from Eq. 6 is used. The black line ($\eta = 0 \,\text{Å}^{-2}$) is equivalent to the plain cutoff function $f_{\text{cut}}(R)$.

Typically, for every combination of central element $I$ and neighboring element $J$, several functions of the type in Eq. 7 with different values of $\eta$ and/or $R_{\text{shift}}$ are used as input features to the NN in Eq. 5. Using several such symmetry functions provides a significantly better fingerprint of the atomic environment than could be accomplished with only a single symmetry function. For example, if the single function with $\eta = 0 \,\text{Å}^{-2}$ in Fig. 3 is used, then two neighbors around the atom $i$, both at a distance $R = 3 \,\text{Å}$, yield the same value of $G^I(i) = 0.197$ as a single neighbor at a distance $R = 2 \,\text{Å}$. By using several symmetry functions, the aim is to "encode" all the relevant information about the atomic environment around an atom $i$ into the input vector $\mathbf{G}^I(i)$ and to provide a structural fingerprint of the atomic environment as the input to the NN.

The radial symmetry function in Eq. 7 is only distance-dependent or in other words spherically symmetric. Therefore it is required to incorporate also angular dependencies into the symmetry functions. A common choice for an angular symmetry function is (Behler 2011a)

$$G^I(i \in I, \{\mathbf{R}, Z\};\ \eta, \zeta, \lambda, J, K)$$

$$= \varphi \left( 2^{1-\zeta} \sum_{\substack{j \in J,\ k \in K \\ j \neq i,\ k \neq i \\ k \neq j}} (1 + \lambda \cos \theta_{jik})^{\zeta} \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} \cdot f_{\mathrm{cut}}(R_{ij}) \cdot f_{\mathrm{cut}}(R_{ik}) \cdot f_{\mathrm{cut}}(R_{jk}) \right)$$

(8)

where the interatomic distances $R_{ij}$, $R_{ik}$, and $R_{jk}$ and angle $\theta_{jik}$ between three atoms $i \in I$, $j \in J$, and $k \in K$ are used to compute the value of the symmetry function, for each of the possible unique combinations of neighbors $j$ and $k$ around the central atom $i$. Again, the elements $J$ and $K$ may be the same as or different from $I$, and $\varphi$ is a scaling function. In Eq. 8, $\zeta$ determines the range of angles for which the angular term is approximately 0, and $\lambda$ takes on a value of either $+1$ or $-1$. The angular part of the symmetry function is shown in Fig. 3b, for a few different values of $\zeta$ and $\lambda$. Notably, the angular part is periodic with a period of $360°$ and symmetric around $0°$ and $\pm 180°$.

By using a cutoff radius, long-range interactions like electrostatic interactions are truncated at the cutoff. This can be problematic for many types of systems. An approach for including long-range interactions into the PES using high-dimensional NNPs is described in Sect. 5.

A high-dimensional NNP can also be used to evaluate the analytic forces acting on atoms, which is central to applications like molecular dynamics simulations. The force with respect to some atomic coordinate $\alpha$ is

$$F_\alpha = -\frac{\partial E}{\partial \alpha} = -\sum_{j=1}^{N^{\mathrm{atom}}} \frac{\partial E_j}{\partial \alpha} = -\sum_{J \in \{Z\}} \sum_{j \in J} \sum_{\mu=1}^{N^{\mathrm{sym}}(J)} \frac{\partial E_j}{\partial G_\mu^J(j)} \cdot \frac{\partial G_\mu^J(j)}{\partial \alpha}$$

(9)

where the outermost sum runs over all chemical elements $J$ in the system and where $G_\mu^J(j)$ is the $\mu$th symmetry function for the element $J$ evaluated for the atom $j \in J$.

## 4    Construction of a High-Dimensional NNP

The construction of a high-dimensional NNP for some given chemical system is a procedure involving

- The procurement of training and validation data
- The choice of symmetry functions (including the cutoff radius $R_{\mathrm{cut}}$) for each element
- The choice of network architecture for each element
- The fitting of the weight matrices $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$ for each element

Moreover, the fitted NNP must be critically evaluated and inspected for regions of the potential energy surface that are not well-described.

In typical applications of high-dimensional NNPs, 1–3 hidden layers and 10–40 nodes per hidden layer are used for each elemental NN. The number of symmetry functions strongly depends on the number of chemical elements in the system. Typically, for an element $I$, 5–10 radial symmetry functions (Eq. 7) per possible neighboring element $J$, and 5–10 angular symmetry functions (Eq. 8) per possible combination of neighboring elements $J$ and $K$, are used. The cutoff distance $R_{\text{cut}}$ is typically set to be in the range from 6 to 10 Å consistently for all symmetry functions.

## 4.1    Procurement of Training Data

The procurement of training data is essential to the successful application of a high-dimensional neural network potential. The neural network can only be as good as the data to which it has been trained. Therefore, it is important to select a reference computational method that is accurate for the modeled material. Often, some form of density functional theory (DFT) is used as the reference method. The training set then consists of a set of structures, with energies that have been determined using DFT. Optionally, the DFT forces acting on the atoms, which contain valuable local information about the PES, can also be used to train the NN. The aim of the training procedure is for the NN to, as closely as possible, reproduce the reference energies and forces; this is achieved by iteratively optimizing the weight matrices $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$ for each elemental NN.

In a typical setting, more and more training data is iteratively added to the training set, in order to obtain better neural network potentials. From some initial collection of training data, several high-dimensional neural network potentials $\chi_0^{(0)}$, $\chi_0^{(1)}$, etc. are trained (as described in the next section). By applying those potentials to, for example, molecular dynamics simulations, it is possible to identify structures that are poorly described. This can be accomplished in several ways:

1. By monitoring the values of the symmetry functions for the different structures that appear in the simulation. If a symmetry function value lies outside the range of values which appear in the training set, or if the value of the symmetry function only appears very rarely in the training set, then the prediction made by the neural network is likely inaccurate.
2. By comparing the energies and forces predicted by different neural networks $\chi_0^{(0)}$, $\chi_0^{(1)}$, etc. on identical structures. If the energies and/or forces are very different for different fits of the neural network, then the respective part of configuration space is insufficiently sampled.

The structures identified in any of above ways are then recalculated using the reference method and added to the training set, which is followed by new NN

optimizations, yielding new NN potentials $\chi_1^{(0)}$, $\chi_1^{(1)}$, etc. In this manner, more and more training data is iteratively added to the training set, until the NN is deemed to accurately describe the PES of all structures that appear in the simulation.

## 4.2   The Choice of Symmetry Functions

The vector of symmetry functions, $\mathbf{G}^I$ in Eq. 5, for each element $I$, is what allows the NNP to discriminate between different structures. Thus, an important consideration, when designing a NNP for some chemical system, is which symmetry functions $G_\mu^I$ to use when evaluating $\mathbf{G}^I$. If only a few symmetry functions are used, the NNP may not reliably be able to distinguish between different structures.

In practice, the set of symmetry functions is often empirically chosen and refined until the NNP gives satisfactory results. A "good" set of symmetry functions normally contains functions fulfilling the following conditions:

- The value of a given symmetry function is not the same for all atoms in the training set. More generally, the range of values for a certain symmetry function when calculated for different atoms in the training set should not be too small, since the NN could then assign large changes in the total energy to small changes of the symmetry function values. In addition, the distribution of values of a certain symmetry function should be analyzed, so as to ensure that the range of values is not dominated by a few outliers.
- The set of symmetry functions covers a range of chemically meaningful distances. For example, the most quickly decaying radial symmetry function (the one with the largest value of $\eta$, Eq. 7) should decay around the distance of the shortest possible meaningful bond between atoms of elements $I$ and $J$.
- Atoms that experience very different forces (as calculated by the reference method) necessarily exist in substantially different environments and should therefore have substantially different values for at least one of the symmetry functions. If this is not the case, then the set of symmetry functions needs to be augmented.
- The correlation between any two symmetry functions should not be too large.

It can be useful to "precondition" input features to the neural network. The optional scaling function $\varphi$ in the example symmetry functions in Eqs. 7 and 8 could, for example, ensure that the symmetry function values for different structures in training set all lie within some predefined range, for example, $[-1, 1]$. The definition of $\varphi_\mu$ would then, for the symmetry function $G_\mu$, be

$$\varphi_\mu(x) = \frac{2(x - G_\mu^{\circ,\min})}{G_\mu^{\circ,\max} - G_\mu^{\circ,\min}} - 1 \qquad (10)$$

where $G_\mu^{\circ,\text{min}}$ and $G_\mu^{\circ,\text{max}}$ are the minimum and maximum values in the training set obtained for the corresponding unscaled symmetry function $G_\mu^\circ$. This kind of feature scaling helps to balance the relative importances of different symmetry functions. It is also possible to use other types of feature scaling functions.

## 4.3    Optimization of the NN Weights

The optimization of the NN weight parameters $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$ for each chemical element starts with an initial guess of their values. The initial weights could simply be random numbers, but there also exist several ways in which the initial weights can be chosen so as to minimize the time needed to train the neural network, for example, the scheme developed by Nguyen and Widrow (1990). It can also be useful to set the weights connecting the last hidden layer with the output layer so that the average and standard deviation of the initially predicted energies (before any training) matches the average and standard deviation of the reference energies in the training set. The reference energies are invariably provided with some unit; the weights can be thought of as dimensionless numbers that give NN energies in the correct unit.

The optimization of the weights is achieved by iteratively minimizing the cost function $\Gamma$. An optimization iteration is often referred to as an epoch. The cost function is frequently taken to be the weighted average of the squared differences between the NN-calculated and reference values. For example, the cost function for the energies, $\Gamma_E$, is

$$\Gamma_E = \frac{1}{N^{\text{struct}}} \sum_{i=1}^{N^{\text{struct}}} w_{i,E} \left( \frac{E_i^{\text{NN}} - E_i^{\text{Ref}}}{N_i^{\text{atom}}} \right)^2 \tag{11}$$

where $N^{\text{struct}}$ is the number of structures in the training set, $w_{i,E}$ is a structure-dependent weight parameter, and $N_i^{\text{atom}}$ is the number of atoms in the $i$th structure. By choosing different values of $w_{i,E}$ for different structures in the training set, it is possible to assign a greater importance to some structures in the training set than to others. Note that the weight $w_{i,E}$ which indicates the importance of the structure $i$ in the training set is set by the user and not fitted, unlike the NN weights $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. In practice, it is common to assign the same weight $w_{i,E} = 1$ to all structures in the training set.

Similarly, the cost function for the forces, $\Gamma_F$, becomes

$$\Gamma_F = \frac{1}{N^{\text{struct}}} \sum_{i=1}^{N^{\text{struct}}} \left[ \frac{w_{i,F}}{3N_i^{\text{atom}}} \sum_{j=1}^{3N_i^{\text{atom}}} (F_j^{\text{NN}} - F_j^{\text{Ref}})^2 \right] \tag{12}$$

where the inner sum is taken over all $3N_i^{\text{atom}}$ force components in the training set structure $i$. If the NN is optimized to both the energies and forces, it is possible to

either alternate between the $\Gamma_E$ and $\Gamma_F$ cost functions for each optimization step (where the NN weights $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$ are updated) or to combine them into a single cost function

$$\Gamma = \Gamma_E + \Gamma_F \tag{13}$$

In this case, the structure-dependent weights $w_{i,F}$ are usually set to much smaller numbers than $w_{i,E}$, since for the structure $i$, there is only a single energy $E_i$ but $3N_i^{\text{atom}}$ force components. The expression for the cost function can also depend on the optimization algorithm; for example, some algorithms include history-dependent terms in the cost function.

Many possible optimization algorithms exist. The simplest one, steepest descent (or "backpropagation" (Rumelhart et al. 1986)), updates the weights according to the following rule:

$$a_{ij}^{(k)}(t+1) = a_{ij}^{(k)}(t) - \eta \cdot \frac{\partial \Gamma}{\partial a_{ij}^{(k)}(t)} \tag{14}$$

where $t$ is the epoch number and $\eta$ is the learning rate. This step is performed for each elemental NN. However, other weight optimization algorithms also exist and are frequently used, such as the Levenberg-Marquardt algorithm (Levenberg 1944; Marquardt 1963) and the global extended Kalman filter (Haykin 2001). The description of those algorithms lies outside the scope of the current chapter.

Often, the quality of a set of NN weights is characterized by the root mean squared error, RMSE, for the energies and forces:

$$\text{RMSE}(E) \text{ per atom} = \sqrt{\frac{1}{N^{\text{struct}}} \sum_{i=1}^{N^{\text{struct}}} \left( \frac{E_i^{\text{NN}} - E_i^{\text{Ref}}}{N_i^{\text{atom}}} \right)^2} \tag{15}$$

$$\text{RMSE}(F) = \sqrt{\frac{1}{N^{\text{struct}}} \sum_{i=1}^{N^{\text{struct}}} \left[ \frac{1}{3N_i^{\text{atom}}} \sum_{j=1}^{3N_i^{\text{atom}}} (F_j^{\text{NN}} - F_j^{\text{Ref}})^2 \right]} \tag{16}$$

$\text{RMSE}(E)$ is often reported as a value normalized "per atom," because structures containing many atoms typically have larger absolute errors in the energy as compared to structures containing few atoms. For $\text{RMSE}(F)$, the inner sum is taken over all $3N_i^{\text{atom}}$ force components of the $i$th structure in the training set.

A good NN will have small values for the RMSE. Typical values reported in the literature are about $\text{RMSE}(E) = 1$ meV per atom and $\text{RMSE}(F) = 100$ meV/Å. However, what constitutes a good fit depends on how the NN will be applied. Moreover, the RMSE provides only a simple measure for how well the trained NN performs on the training data. If the training data is very varied, for example, with respect to the chemical composition of the system, the RMSE averaged over all the

structures in the training set does not necessarily help the scientist find structures or chemical compositions that are not well described by the NN. In such cases, it is a good idea to split the training set into several different groups, e.g., according to composition and calculate the RMSE for each group separately. In addition, it is helpful to explore the distribution of errors in the energies and forces made by the NN. If the NN performs poorly for some particular set of structures, it can be helpful to increase their weights $w_E$ and $w_F$ in the training procedure (Eqs. 11 and 12) or to modify the set of symmetry functions so that the atomic chemical environments are better described.

## 4.4    Training Set, Validation Set, Test Set, Overfitting and Underfitting

The procured reference data is typically divided into a *training set* and a *test set*. The training set consists of the structures used to determine the NN weight parameters. The test set consists of additional reference data that is not used to train the NN. By using a test set, the quality of the fitted NN can be evaluated on structures to which it has not been trained by monitoring the cost function of both sets.

The usage of a test set is instrumental for detecting the so-called overfitting (also known as high variance). An example of overfitting is illustrated in Fig. 4a in a simple one-dimensional case: Although the NN represented by the blue line reproduces the reference values at each training point (red circles) very well, the NN makes quite inaccurate predictions for values that lie outside the training set (orange triangles). Therefore, although the RMSE on the training set would be small, the RMSE for a test set would be considerably greater. This indicates that the NN makes poor interpolations between the data points in the training set. Overfitting normally happens as a result of the neural network architecture being too large, e.g., with respect to the number of hidden layers or the number of nodes per hidden layer. The overfitting can be combated by reducing the size of the neural network, by stopping
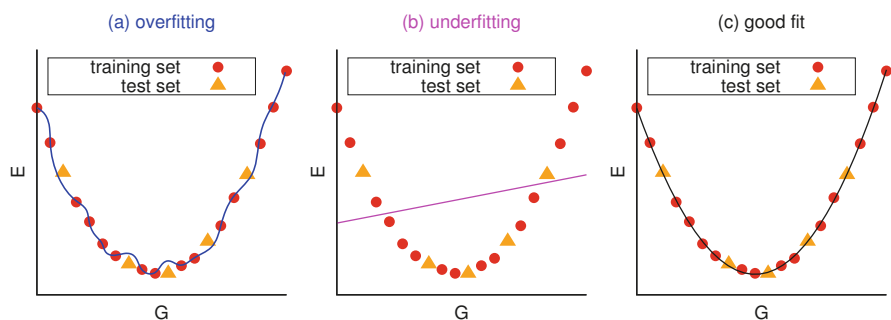


**Fig. 4** Examples of (**a**) overfitting, (**b**) underfitting, and (**c**) a good fit. The red circles represent data points in the training set; the orange triangles represent points in the test set

the NN weight optimization at an earlier epoch, or by including a regularization term to the cost function $\Gamma$. A regularization term penalizes large values of the weights $a_{ij}^{(k)}$, and many types of regularization terms are possible (e.g., the so-called $L_1$ and $L_2$ regularization). A more in-depth discussion on regularization in connection with neural networks can be found in, for example, Haykin (2011).

The opposite of overfitting is underfitting (also known has high bias), illustrated in Fig. 4b. Underfitting is characterized by a high RMSE for both the training and test sets and typically occurs if not enough epochs are run for the NN weight optimization. If running more epochs does not address the underfitting, it can help to make the neural network functional form more flexible, for example, by adding more nodes to the hidden layers.

Figure 4c illustrates a "good fit," where the error between the NN prediction and the reference data is small, as well as roughly equal for both the training and test sets.

If several neural network architectures are evaluated for the purpose of addressing, for example, overfitting, then there is a danger that the finally selected neural network architecture has a deceivingly small error on the test set, since the test set was used to select the neural network architecture. In such cases, the "test set" used to evaluate the NN architecture is commonly referred to as a *validation set*, and the error of the finally selected NN is then evaluated for another, independent, test set.

## 5      Inclusion of Long-Range Electrostatics

The neural network potential described in Sect. 3 relies on a vector of symmetry function values as input, and for any given atom, only atoms in the environment (within the sphere of the cutoff radius $R_{\text{cut}}$) contribute to the symmetry function values. Thus, a fundamental assumption of the type of high-dimensional NNP described in Sect. 3 is that the potential energy of the system can be calculated from local atomic environments. Some type of interactions, in particular electrostatic interactions, decay only slowly with increasing distance and can have considerable impact on energies and forces even if the distance between two interacting atoms is larger than $R_{\text{cut}}$.

The calculation of electrostatic contributions requires information about the charge distribution. This can be achieved by, for example, associating each atom with a point charge. The signs and magnitudes of those charges can either be fixed at some predetermined (element-dependent) values or be determined "on-the-fly" in some manner. One example of the latter type of approach that has been used in conjunction with NNPs, is to use a second, different, NNP (Artrith et al. 2011; Morawietz et al. 2012). This second NNP would then be fitted to reproduce the atomic charges, akin to how the type of NNP previously introduced is used to determine atomic energies. With this approach, some approximation for the atomic charges must be provided in the training set. Typically, one of the many types of atomic charges that can be derived from electronic structure calculations, such

as Mulliken charges (Mulliken 1955), Bader charges (Bader 1985), or Hirshfeld charges (Hirshfeld 1977), is used. Figure 5 illustrates this approach, where one high-dimensional NNP is used to compute the "short-range" energy $E^{short}$ that only depends on the local atomic environments, and a second high-dimensional NNP is used to estimate the atomic charges that are used to compute the "long-range" energy $E^{long}$ (via, e.g., an application of Coulomb's law or Ewald summation). The potential energy is then obtained as $E = E^{short} + E^{long}$. The NN architectures and symmetry functions used for the different elements to calculate atomic charges need not be the same as those used to calculate the atomic energies.

Regardless of the scheme used to generate the atomic charges and to calculate $E^{long}$, it is crucial that the short-range NN is fitted to reproduce only the *difference* between $E^{tot}$ and $E^{long}$, in order to avoid double-counting of contributions to the total energy. In the calculation of $E^{long}$, it has been shown to be beneficial to use a Coulomb potential that is screened at short distances, since otherwise the NN might need to fit a more corrugated potential energy surface for the short-range energies.

The force with respect to some atomic coordinate $\alpha$ becomes

$$F_\alpha = F_\alpha^{short} + F_\alpha^{long} = -\frac{\partial E^{short}}{\partial \alpha} - \frac{\partial E^{long}}{\partial \alpha} \tag{17}$$

where $F_\alpha^{short}$ is calculated as in Eq. 9. For a nonperiodic system, $F_\alpha^{long}$ can be calculated as

$$F_\alpha^{long} = -\frac{1}{2}\frac{\partial}{\partial \alpha} \sum_{i=1}^{N^{atom}} \sum_{\substack{j=1 \\ j \neq i}}^{N^{atom}}$$

$$\frac{q_i q_j}{R_{ij}} = -\frac{1}{2} \sum_{i=1}^{N^{atom}} \sum_{\substack{j=1 \\ j \neq i}}^{N^{atom}} \frac{1}{R_{ij}^2} \left[ \frac{\partial q_i}{\partial \alpha} q_j R_{ij} + q_i \frac{\partial q_j}{\partial \alpha} R_{ij} - q_i q_j \frac{\partial R_{ij}}{\partial \alpha} \right] \tag{18}$$

where $q_i$ is the charge on atom $i$. If a NNP with symmetry functions is used to determine the atomic charges, as in Fig. 5, then it can be shown that

$$F_\alpha^{long} = \sum_{J \in \{Z\}} \sum_{j \in J} \sum_{\substack{i=1 \\ i \neq j}}^{N^{atom}} \frac{q_i}{R_{ij}} \cdot \left[ \frac{1}{2}\frac{q_j}{R_{ij}}\frac{\partial R_{ij}}{\partial \alpha} - \sum_{\mu=1}^{N^{sym}(J)} \frac{\partial q_j}{\partial G_\mu^J(j)}\frac{\partial G_\mu^J(j)}{\partial \alpha} \right]. \tag{19}$$

For a periodic system, where the electrostatic energy is evaluated using, for example, Ewald summation, the expression for $F_\alpha^{long}$ becomes more complicated, although it can be derived in a similar fashion.
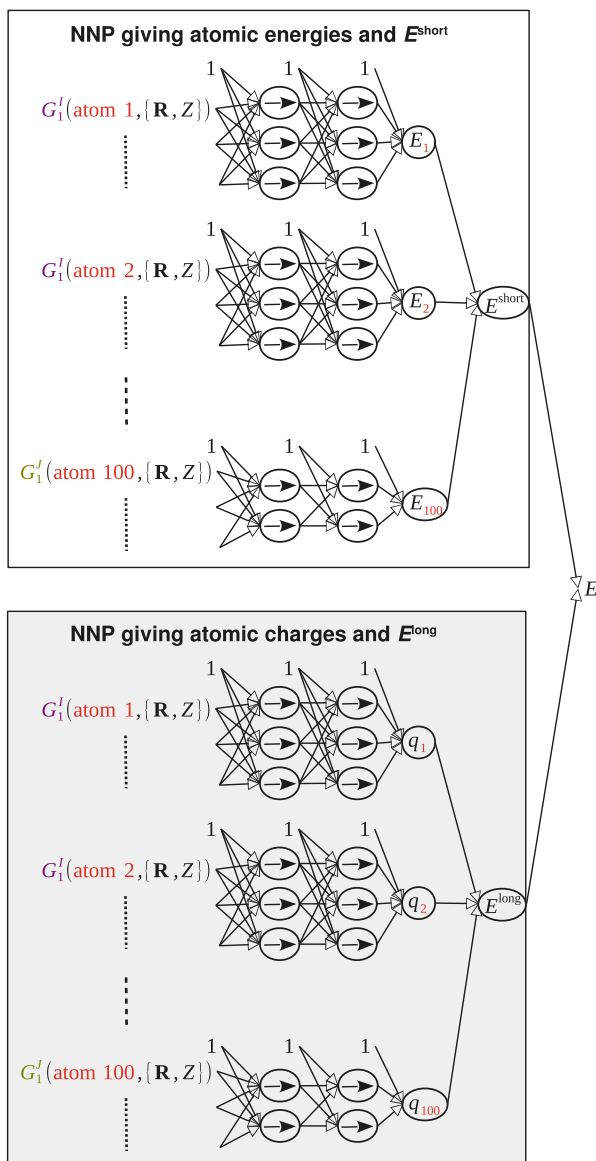
**Fig. 5** Illustration of two high-dimensional neural network potentials, using the same chemical system and color coding as in Fig. 2. The upper NN (white background) yields the short-range energy $E^{\text{short}}$; the lower NN (gray background) yields the atomic charges $q_i$ that are used to compute $E^{\text{long}}$. Together they give the total energy $E$

# 6    Applications and Limitations of Neural Network Potentials

NNPs, and in particular high-dimensional NNPs, have been developed and applied to a range of different molecules and materials (for reviews see Behler 2014, 2017). Some examples include silicon (Behler et al. 2008), carbon (Khaliullin et al. 2011), sodium (Eshet et al. 2012), zinc oxide (Artrith et al. 2011), germanium telluride (Sosso et al. 2012), copper (Artrith and Behler 2012), Cu clusters on ZnO (Artrith et al. 2013), Cu-Au nanoalloys (Artrith and Kolpak 2015), water on Cu (Natarajan and Behler 2016), titanium dioxide (Artrith and Urban 2016), gold (Boes et al. 2016), copper-palladium-silver alloys (Hajinazar et al. 2017), $N_2$ on Ru (Shakouri et al. 2017), and water on ZnO (Quaranta et al. 2017).

Because of the unprecedented accuracy that can be obtained at low computational cost with such a flexible method, the development of NNPs, and of the NNP methodology, but also of other similar ML potential methods, is currently a very active research field. Some limitations and drawbacks of high-dimensional NNPs, as they have been presented in this chapter, include the following:

- *The iterative construction of large training sets*. Typical training sets include thousands of structures that must be computed using a reference electronic structure method. Moreover, the generation of new training set structures is often done in an empirical fashion, using, for example, high-temperature molecular dynamics simulations.
- *Limited number of elements in the chemical system*. With more than three or four chemical elements in the system, the number of needed symmetry functions to describe the local chemical environment around an atom becomes exceedingly large. For such systems, another type of input feature might be beneficial.
- *Complicated and time-consuming fitting procedure*. NNPs contain many weights that need to be fitted. Even with the use of advanced fitting algorithms, obtaining a good fit is often a time-consuming procedure. This problem is exacerbated with the inclusion of long-range electrostatics as outlined in Sect. 5.

# 7    Summary

In recent years, neural network potentials (NNPs) have become a very useful tool for atomistic materials modeling simulations. They provide the potential energy surface (PES) of a system at low computational cost, while the accuracy is very close to that of first-principles methods. Because of the flexible functional form of NNPs, they can be used to describe all types of atomic interactions (covalent bonding, dispersion interactions, hydrogen-bonding, etc.) on an equal footing. In case of significant long-range electrostatic interactions, NNPs can be extended with some scheme for determining reference atomic charges or multipoles, to include the long-range contribution to the total energy. While NNPs are fitted to reproduce reference data (typically total energies and atomic forces) from electronic structure

calculations, a careful validation of the obtained potentials is required due to the intrinsically nonphysical but purely mathematical functional form. In case of high-dimensional NNPs, one NN is fitted for each chemical element in the system. The element-specific NN then describes the PES around a particular type of atom. The local atomic environment is transformed into input for the NN using, for example, symmetry functions. The total energy is then calculated by summing over all atomic contributions.

The errors associated with NNPs, as compared to the reference electronic structure method, can be made extremely small ($< 1$ meV per atom), which makes NNPs a promising tool for future applications in materials modeling and simulation.

# References

Artrith N, Behler J (2012) High-dimensional neural network potentials for metal surfaces: a prototype study for copper. Phys Rev B 85:045439

Artrith N, Kolpak AM (2015) Grand canonical molecular dynamics simulations of Cu–Au nanoalloys in thermal equilibrium using reactive ANN potentials. Comput Mater Sci 110:20

Artrith N, Urban A (2016) An implementation of artificial neural-network potentials for atomistic materials simulations: performance for $TiO_2$. Comput Mater Sci 114:135–150

Artrith N, Morawietz T, Behler J (2011) High-dimensional neural-network potentials for multi-component systems: applications to zinc oxide. Phys Rev B 83:153101

Artrith N, Hiller B, Behler J (2013) Neural network potentials for metals and oxides – first applications to copper clusters at zinc oxide. Phys Status Solidi B 250:1191–1203

Bader R (1985) Atoms in molecules. Acc Chem Res 18:9

Balabin RM, Lomakina EI (2011) Support vector machine regression (LS-SVM)-an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? Phys Chem Chem Phys 13:11710

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403

Behler J (2011a) Atom-centered symmetry functions for constructing high-dimensional neural network potentials. J Chem Phys 134:074106

Behler J (2011b) Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. Phys Chem Chem Phys 13:17930–17955

Behler J (2014) Representing potential energy surfaces by high-dimensional neural network potentials. J Phys Condens Matter 26:183001

Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145(17):170901

Behler J (2017) First principles neural network potentials for reactive simulations of large molecular and condensed systems. Angew Chem Int Ed 56:12828

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98:146401

Behler J, Martoňák R, Donadio D, Parrinello M (2008) Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. Phys Rev Lett 100:185501

Bishop CM (1996) Neural networks for pattern recognition. Oxford University Press, Oxford

Blank TB, Brown SD, Calhoun AW, Doren DJ (1995) Neural network models of potential energy surfaces. J Chem Phys 103:4129–4137

Boes JR, Groenenboom MC, Keith JA, Kitchin JR (2016) Neural network and ReaxFF comparison for Au properties. Int J Quantum Chem 116:979–987

Eshet H, Khaliullin RZ, Kühne TD, Behler J, Parrinello M (2012) Microscopic origins of the anomalous melting behavior of sodium under high pressure. Phys Rev Lett 108:115701

Hajinazar S, Shao J, Kolmogorov AN (2017) Stratified construction of neural network-based interatomic models for multicomponent materials. Phys Rev B 95:014114

Handley CM, Popelier PLA (2010) Potential energy surfaces fitted by artificial neural networks. J Phys Chem A 114:3371–3383

Haykin S (2001) Kalman filtering and neural networks. Wiley, Hoboken

Haykin S (2011) Neural networks and learning machines. Pearson Education, New Dehli

Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theor Chim Acta 44:129–138

Khaliullin RZ, Eshet H, Kühne TD, Behler J, Parrinello M (2011) Nucleation mechanism for the direct graphite-to-diamond phase transition. Nat Mater 10:693–697

Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. Quart Appl Math 2:164–168

Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J Appl Math 11:431–441

Morawietz T, Sharma V, Behler J (2012) A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. J Chem Phys 136:064103

Mulliken RS (1955) Electronic population analysis on LCAO-MO molecular wave functions. I. J Chem Phys 23:1833

Natarajan SK, Behler J (2016) Neural network molecular dynamics simulations of solid-liquid interfaces: water at low-index copper surfaces. Phys Chem Chem Phys 18:28704

Nguyen DH, Widrow B (1990) Neural networks for self-learning control systems. IEEE Control Syst Mag 3:18–23

Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, Oxford

Quaranta V, Hellström M, Behler J (2017) Proton transfer mechanisms at the water-ZnO interface: the role of presolvation. J Phys Chem Lett 8:1476

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536

Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA (2012) Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 108:058301

Shakouri K, Behler J, Meyer J, Kroes GJ (2017) Accurate neural network description of surface phonons in reactive gas-surface dynamics: $N_2$+Ru(0001). J Phys Chem Lett 8:2131

Sosso GC, Miceli G, Caravati S, Behler J, Bernasconi M (2012) Neural network interatomic potential for the phase change material GeTe. Phys Rev B 85:174103

# Part V
# Long-Timescale Atomistic Simulations: Accelerated Molecular Dynamics and Adaptive Kinetic Monte Carlo

# Computational Methods for Long-Timescale Atomistic Simulations

# 32

Blas Pedro Uberuaga and Danny Perez

## Contents

**Abstract**

Because of their considerable predictive power, atomistic simulations are extremely powerful tools in the computational materials scientist's toolbox. This power however comes at a significant computational price that rather strongly limits the accessible simulation space, especially in terms of the timescales that can be directly simulated. Specialized methods specifically designed to overcome timescale limitations while still faithful to the underlying dynamical behavior of the system are hence essential in order to bridge the gap with experiments. This section summarizes the most recent advances in a class of open-ended long-timescale atomistic simulation techniques that include accelerated molecular dynamics and kinetic Monte Carlo methods. The different chapters introduce the basics of these methods as well as a review of their most recent developments.

B. P. Uberuaga (✉)
Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: blas@lanl.gov

D. Perez
Theoretical Division T-1, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: danny_perez@lanl.gov

# 1 Introduction: The Timescale Limitations of Molecular Dynamics

Molecular dynamics simulations have become a workhorse in the general endeavor of computational materials science. By providing an atomic-scale description of materials, each and every atom can be followed and its behavior tracked computed, a resolution simply not possible with any other technique, experimental or otherwise. These types of simulations have provided unprecedented insight into the behavior of materials. For example, some of the first molecular dynamics simulations were on the problem of collision cascades during irradiation. Indeed, it is from such simulations that the majority of our knowledge about the primary damage state created during a radiation damage event arises. Similarly, knowledge of the complex behavior of dislocations, interfaces, or grain boundaries is in great part due to our ability to carry out direct atomistic simulations.

However, despite the immense popularity of molecular dynamics in materials research, significant shortcomings still limit the types of studies that can be performed. Chief among these is the limited timescales. While heroic efforts can reach microseconds for some systems, most are limited to nanoseconds, and all fall far short of the timescales relevant to most experimental conditions. Coming back to the problem of radiation damage, one would ideally wish to predict the response of the material over years of irradiation, or at least be able to simulate on timescales that correspond to collision cascades hitting the same region of space more than once at realistic fluxes. The art of molecular dynamics therefore often lies in carefully designing simulation protocols or initial conditions such that long-time behavior can be inferred from short simulations. While this is sometimes possible, it is often extremely challenging to do accurately, which severely limits the contribution of molecular dynamics to answering key questions in materials science.

It is important to note that this limitation persists regardless of the size of the computer that is employed. This is due to the intrinsically serial nature of the numerical solution of the atomistic equations of motion: work on a given timestep has to be completed before work on the next can begin. Therefore, effective parallelization of conventional molecular dynamics can only proceed within a timestep, not across. This usually translates into a spatial parallelization scheme where different processing elements are in charge of computing the forces acting on atoms in distinct physical regions, an approach that is efficient only if every computing element is assigned enough work that communication or synchronization costs are negligible; it however breaks down when the number of atoms assigned to each processing element becomes too small. Parallelization of conventional molecular dynamics is therefore a viable approach to extend the simulation length-scales, but typically not their timescales. Thus, there has been a recognized need for methods that retain the full atomistic fidelity of the underlying interatomic model but push the accessible timescale to those more amenable to direct comparison with experiment.

## 2 Accelerating Rare Event Dynamics

With this motivation in mind, a number of scientists have developed fundamentally new simulation approaches with precisely this goal. Such efforts have traditionally been focused on systems that evolve through so-called rare events, where long waiting periods, often corresponding only to vibrational motion, are punctuated by rare but rapid transitions to different configurations. This class of materials is especially crippled by the limitations of standard approaches, as short simulations are extremely unlikely to be informative of their long-time behavior. The exact nature of what makes these events rare can vary between materials and external conditions, ranging from high energetic barriers to entropic constrictions. These obstructions lead to a so-called separation of timescales between fast (corresponding to local exploration of the potential energy surface) and slow (corresponding to rare transitions over the kinetic bottlenecks) components of the dynamics. This separation enables a statistical treatment of the slow events, as local equilibrium with respect to the fast degrees of freedom can be assumed. The simplifications that follow from this assumption are key to enabling most of the long-timescale methods.

Efforts have long been underway in the development of techniques that can exploit the separation of timescales to extend the predictive power of atomistic simulations beyond what can directly be simulated. In broad terms, these techniques can be classified into what we term *kinetic* and *dynamic* methods. Kinetic techniques are designed to compute transition rates or sample from the ensemble of reactive trajectories for rare transitions that occur along pre-specified reaction pathways or that otherwise connect specific regions of configuration space (e.g., the solid and liquid regions). This family contains a very large number of methods, including free energy (see ▶ Chaps. 29, "Variationally Enhanced Sampling" and ▶ 27, "Metadynamics: A Unified Framework for Accelerating Rare Events and Sampling Thermodynamics and Kinetics"), path sampling, and milestone-based approaches. These tools are especially powerful when one is interested in characterizing a process which is known to occur or when the basic nature of phenomena of interest is known a priori, e.g., when one can identify or construct a proper reaction coordinate for the process of interest. These methods are also adept at handling energy landscapes that are largely smooth on energy scales larger than $k_BT$, hence their popularity in the soft matter community. When this is not the case, truly dynamical methods, where a single (or a small number of) unbiased trajectories can be generated – just like in standard molecular dynamics – are preferred. This family of techniques is in comparison much smaller.

## 3 Summary of the Section

In order to allow for a comprehensive overview, we here restrict our attention to the second class of methods, and especially to accelerated molecular dynamics and adaptive kinetic Monte Carlo techniques. These methods are often very effective

for hard materials at moderate temperatures, but they tend to fare less well for softer systems where the topology of the fast-equilibrating regions of configuration space can be very complex. For these systems, kinetic methods, where additional knowledge can be exploited, are often preferred. The most recent developments in dynamical methods are described in this section, written by global leaders in the field. These approaches broadly fall into two categories: (i) bypass the problem altogether by foregoing the integration of the equations of motion in favor of direct searches for escape pathways or (ii) use molecular dynamics as a computational engine, but "trick" it into revealing relevant kinetic information faster.

The first approach requires methods to search and characterize the potential energy surface of the system. Hannes Jónsson is the original developer of the nudged elastic band (NEB) method (Jónsson et al. 1998), a method for finding saddle points on potential energy surfaces given an initial and final state of the system. He and his colleague describe recent developments in such static approaches to finding saddle points and to computing transition rates along these pathways, for example, using transition state theory (TST) (see ▸ Chap. 33, "Exploring Potential Energy Surfaces with Saddle Point Searches"). Given the static nature of these types of calculations, they are more amenable to application with electronic structure approaches such as density functional theory than are dynamics-based methods. While these methods can be characterized as kinetic according to the above discussion, they are instrumental in the design of truly dynamical methods like adaptive kinetic Monte Carlo (KMC) (Henkelman and Jónsson 2001). KMC techniques can generate long trajectories that overcome many of these bottlenecks without having to assume that their nature or kinetics is known a priori. The most recent developments in these types of approaches are described by Normand Mousseau and Graeme Henkelman, two pioneers in this field (see ▸ Chap. 34, "Off-Lattice Kinetic Monte Carlo Methods").

The alternative approach is embodied by the so-called accelerated molecular dynamics (AMD) methods (Voter 1997, 1998; Sorensen and Voter 2000), now over 20 years old. Generally, though not exclusively, based on TST, these approaches sacrifice fidelity on the scale of atomic vibration to describe accurate state-to-state dynamics on timescales longer, in some cases much longer, than molecular dynamics. While some of the AMD methods provide acceleration even in serial, the increased availability of massively parallel platforms has provided unique opportunities to further enhance their reach. ▸ Chapter 35, "Accelerated Molecular Dynamics Methods in a Massively Parallel World" by Richard Zamora, Arthur Voter, and co-workers describes a number of such parallelization strategies that can be employed to extend the temporal reach of molecular dynamics.

Tony Lelièvre delves into recent developments in the fundamentals of the AMD methods, putting them on a solid mathematical foundation (see ▸ Chap. 36, "Mathematical Foundations of Accelerated Molecular Dynamics Methods"). Many of these new insights stem from the crucial concept of quasi-stationary distribution, which has proved extremely powerful to understand the behavior of systems that evolve through rare events. This concept can be used to reinterpret all of the AMD methods and gain new insights on the precise nature of the assumptions that they

each entail and on their relative strengths and limitations. In particular, this formal framework has been used to show that one method, parallel replica dynamics, can be made arbitrarily accurate by adjusting a single free parameter (Le Bris et al. 2012). This chapter also casts the other traditional AMD methods in this new framework.

While AMD and adaptive KMC methods are adept at reaching long timescales on relatively small systems, simultaneously increasing the timescales and length-scales is an ongoing challenge. Ellad Tadmor and Woo Kyun Kim show how concurrent multiscale methods that reduce the number of degrees of freedom that need to be explicitly simulated for systems where capturing the proper elastic response of the environment is crucial (such as the quasicontinuum method (Shenoy et al. 1999)) can be coupled with the AMD methods to yield extended simulation times while preserving the fidelity of the simulation (see ▶ Chap. 37, "Temporal Acceleration in Coupled Continuum-Atomistic Methods").

Finally, Arthur Voter, Tony Lelièvre, Graeme Henkelman, Normand Mousseau, and Hannes Jónsson all come together for a joint chapter on the challenges and best practices associated with the development and use of long-timescale methods (see ▶ Chap. 38, "Long-Timescale Simulations: Challenges, Pitfalls, Best Practices, for Development and Applications"). Broken into four questions that get at the heart of these approaches, each offers his perspective on the state of the art, particular pitfalls that need to be considered when developing and implementing long-timescale methods, and challenges in using them. They also discuss where they think the future of these methods lies.

Together, these chapters provide both a brief introduction but, more importantly, a current perspective on the state of the art of long-timescale atomistic simulations. Each chapter highlights the current capability of each method as well as potential opportunities for next steps in advancing the method and the field. With the advent of massively parallel computing architectures, these approaches will find new application that will extend our collective knowledge of the complex atomic-scale behavior that underlies much of the macroscopic materials properties that scientists worldwide are investigating. Thus, these methods will become even more central in the ongoing quest to understand materials.

## References

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115:9657–9666

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method. In: Berne BJ, Ciccotti G, Coker DF (eds) Classical and quantum dynamics in condensed matter phase simulations. World Scientific, Singapore. pp 385–404

Le Bris C, Lelièvre T, Luskin M, Perez D (2012) A mathematical formalization of the parallel replica dynamics. Monte Carlo Methods Appl 18:119–146

Shenoy VB, Miller R, Tadmor EB, Rodney D, Phillips R, Ortiz M (1999) An adaptive finite element approach to atomic-scale mechanics – the quasicontinuum method. J Mech Phys Solids 47: 611–642

Sorensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112:9599–9606

Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett 78:3908

Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57:R13985

# Exploring Potential Energy Surfaces with Saddle Point Searches

## Vilhjálmur Ásgeirsson  and Hannes Jónsson

## Contents

V. Ásgeirsson
Science Institute of the University of Iceland, Reykjavík, Iceland
e-mail: via9@hi.is

H. Jónsson (✉)
Faculty of Physical Sciences, University of Iceland, Reykjavík, Iceland

Department of Energy Conversion and Storage, Technical University of Denmark, Lyngby, Denmark
e-mail: hj@hi.is

**Abstract**

The energy surface of an atomic scale representation of a material contains the essential information needed to determine the structure and time evolution of the system at a given temperature. Local minima on the surface represent (meta)stable states of the system, while first-order saddle points characterize the mechanisms of transitions between states. While many well-known methods make it relatively easy to find local minima, the identification of saddle points is more challenging. In this chapter, methods for finding saddle points are discussed as well as applications to materials simulations. Both doubly constrained search methods, where the final and the initial state minima are specified, and singly constrained search methods, where only the initial state is specified, are discussed. The focus is on a classical description of the atom coordinates, but saddle points corresponding to quantum mechanical tunneling are also mentioned. An extension to magnetic systems where the energy surface depends on the orientation of the magnetic vectors is sketched.

# 1     Introduction

Atomic scale simulations of materials often involve finding likely arrangements of the atoms as well as identifying the mechanism and estimating the rate of transitions between different arrangements such as diffusion events, migration of dislocations, chemical reactions, and phase transformations. These can be obtained by analyzing the energy surface characterizing the system, i.e., the electronic ground-state energy of the system as a function of the location of the atoms (and possibly also the orientation of the moments of magnetic atoms) (Peters 2017). Each state corresponds to a minimum on the energy surface and the potential well surrounding it. Excited electronic states can, in some circumstances, be important, but the discussion here will focus on the electronic ground state. Even so, the energy surface is complex and multidimensional for most material systems of interest, and it requires efficient tools to navigate on the energy surface and extract the desired information.

Given some initial configuration of the atoms and a method for evaluating the energy and its gradient with respect to atom coordinates, the nearest local minimum can readily be identified with any number of available minimization methods. A significant challenge, however, is to find all the relevant minima for given conditions, such as temperature. The lower the local energy minimum of the state, the more likely the system is to be in that state, according to Boltzmann statistics. The width of the energy well around the minimum and the number of equivalent minima relate to the vibrational and configurational entropy, respectively, and also contribute to the stability of the state. The identification of the most likely state of a system requires navigation on the energy surface, for example, advancing from one energy minimum to another.

One way to discover new minima is to move through regions of dips in the energy ridge, i.e., regions around first-order saddle points on the energy surface. At a first-

order saddle point, the gradient of the energy vanishes, and the Hessian, the matrix of second derivatives, has one and only one negative eigenvalue. The eigenvectors are referred to as modes, and the one corresponding to the negative eigenvalue is the unstable mode. A general feature of energy surfaces for atomic systems is that a lower energy saddle point tends to lead to a lower-energy minimum on the other side (the basis of so-called Brønsted-Evans-Polanyi relation). So, more often than not, a low-energy minimum can be discovered by moving over the energy ridge through a low-energy saddle point. Methods for finding first-order saddle points can, therefore, be used to identify local minima and likely states of the system.

It may, however, not be possible to reach some states, such as the one corresponding to the global energy minimum, on a given time scale because of a large energy barrier. The rate at which the system moves from one state to another is the key quantity that is needed to predict the long time scale evolution of the system. On a given time scale, only transitions that are frequent enough are relevant. Classical dynamics simulations based on Newton's equation of motion can be carried out, but the time scale of such simulations is limited by the vibrational frequency of the atoms and even for the simplest description of the energy surface – an empirical potential function – such a simulation can only span a small fraction of a second, while the challenge is often to predict the evolution of a material over years. As is explained in the following section, the first-order saddle points on the energy surface and their vicinity can be used to estimate the rate of transitions between states of the system. The central focus of this chapter is methods for finding first-order saddle points for the purpose of exploring energy surfaces of material systems.

## 2      Estimation of Transition Rates

The dynamics of atoms in materials that are relatively stable under conditions of interest can be characterized by a large number of vibrations back and forth, mostly in the close vicinity of the local energy minimum but occasionally involving large excursions caused by rare fluctuations through coupling to the heat bath. Such fluctuations can be large enough for the system to enter a new state. Since the system spends a long time in each state, a Boltzmann distribution of energy is typically established in all degrees of freedom, and the rate at which the system can transform from one state to another can be estimated to a first approximation by using statistical mechanics.

The basic tool for estimating transition rates in materials is transition state theory (TST) (Wigner 1938; Peters 2017). Assuming the atoms can be described as classical particles and that Boltzmann distribution of energy has been established in all degrees of freedom, the lifetime of a given initial state and possible final states of transitions from it can be estimated in a two-step procedure referred to as Wigner-Keck-Eyring (WKE) dynamics (Wigner 1938; Keck 1967; Eyring 1935). In the first step, a dividing surface in configuration space separating the initial state from other states is defined. The dividing surface should be placed in such a way that it lies through regions of the energy surface where the system is least

likely to be found. The dimensionality of the dividing surface is $D - 1$ where $D$ is the total number of degrees of freedom in the system (in three dimensions, $D = 3N$ where $N$ is the number of atoms). By adding a small width around the dividing surface, a subspace of dimensionality $D$ is generated, and this is referred to as the transition state. Transition state theory assumes that if the system makes it to the transition state and has velocity pointing away from the initial state, a transition will occur and the system enter a new state. This basic TST assumption is an approximation. Dynamical trajectories can be reflected back after crossing the transition state and reenter the initial state. They can, in fact, recross multiple times, and each recrossing to the initial state leads to an overestimate of the transition rate in the TST approximation. As a results, TST is guaranteed to give an upper bound to the transition probability. This gives a variational principle for the placement of the dividing surface. The dividing surface that leads to the lowest TST transition rate estimate is the optimal one. The first step of the WKE procedure should, therefore, involve a variational optimization of the dividing surface in addition to giving an approximation to the transition rate, $k^{\text{TST}}$.

The second step of the WKE procedure involves the calculation of short time scale trajectories started from the dividing surface. Since the transition state represents a region of low probability, the system advances quickly from there to either the initial state or some final state. The trajectories in the second WKE step reveal possible final states of the transition. Note that TST does not specify the final state. The trajectories also serve as a means to correct the TST rate estimate and obtain (in principle) the exact estimate of the rate by evaluating a dynamical correction factor, $\kappa$ (Keck 1967; Voter and Doll 1985). The less the dividing surface was optimized in the first step, the larger the number of trajectories needed in the second step to obtain a statistically converged correction factor. A recrossing of the transition state can occur either because of the shape of the energy surface, such as a curved valley on the energy surface in the final state, or because of a fluctuation from the heat bath resulting in a reversal of the relevant component of the velocity. Both effects are included in the short dynamical trajectories of the second WKE step. The WKE procedure, therefore, provides essentially an exact estimate of the transition rate, $k = \kappa k^{\text{TST}}$. Instead of involving impossibly long dynamical trajectories started near the initial state energy minimum, WKE uses a statistical estimate of the probability of making it to the transition state and then short time scale trajectories for the evolution from the transition state to possible final states.

The full implementation of the WKE procedure is in general challenging because of the need to represent and optimize a high-dimensional dividing surface. Efficient tools for such calculations have not been developed yet. The simplest approach is to use a hyperplanar dividing surface where both the location and the orientation are variationally optimized (Jóhannesson and Jónsson 2001). The orientation specifies which atoms are displaced and by how much at the transition state. The orientational optimization identifies the optimal transition mechanism. A single hyperplane is, however, in general not sufficient to specify the full dividing surface, and either a curved surface (Ciccotti et al. 1995) or a mosaic of hyperplanar segments is needed (Bligaard and Jónsson 2005).

The most commonly used form of TST involves an approximation to the energy surface. By expanding the energy to second order in the vibrational normal modes at both the initial state minimum and at a first-order saddle points on the energy rim surrounding the initial state, an approximation to the rate constant is obtained and referred to as harmonic TST (HTST) (Vineyard 1957):

$$k^{\text{HTST}} = \frac{\prod_j^D \nu_{i,\text{m}}}{\prod_j^{D-1} \nu_{i,\text{s}}} \ \exp\left(-(E_\text{s} - E_\text{m})/k_\text{B}T\right) \tag{1}$$

where $E_\text{m}$ and $E_\text{s}$ are the values of the energy the $\nu_{i,\text{m}}$ and $\nu_{i,\text{s}}$ are the vibrational frequencies at the minimum and at the saddle point, $k_\text{B}$ is the Boltzmann constant, and $T$ is the temperature. This expression agrees with the empirical Arrhenius dependence of the rate constant on temperature:

$$k = A \exp\left(-E_\text{a}/k_\text{B}T\right) \tag{2}$$

showing that the activation energy, $E_\text{a}$, is the energy difference between the first-order saddle point and the initial state minimum and the prefactor is related to the width of the potential energy well at the minimum and the energy valley at the saddle point.

The vibrational frequencies are obtained from the positive eigenvalues of the Hessian. The number of vibrational modes at the saddle point is one less than at the minimum because the dividing surface does not contain the unstable mode, the vibrational mode corresponding to negative eigenvalue. The ratio of the products of vibrational frequencies represents the relative vibrational entropy of the transition state and the initial state. Effectively, the dividing surface is approximated in HTST by a hyperplane going through each one of the first-order saddle points on the energy rim surrounding the initial state minimum with normal vectors pointing in the direction of unstable modes. Strictly speaking these should be finite hyperplanar segments, but the full hyperplane is included for each saddle point in order to obtain a simple analytical expression for the rate constant.

The HTST approximation can give an accurate estimate of the rate constant for materials at not too high or too low temperature. For low enough temperature, nuclear quantum effects need to be accounted for (see below). The energy surface must be smooth enough and the first-order saddle points separated by regions of high enough energy for the harmonic approximation to accurately represent the energy surface in the neighborhood of the extrema where the Boltzmann statistical weight is appreciable. This means that the first-order saddle points on the energy ridge must be separated by second-order saddle points that are significantly higher in energy, by several $k_\text{B}T$, compared to the first-order saddle points. Below, we will describe a method that can be used to check whether the energy landscape satisfies this criterion.

The main challenge is to find the relevant saddle points. The lower the energy of a saddle point, the more probable the corresponding transition is. While the number

of saddle points surrounding a given minimum on the energy surface can easily be enormous, it is enough to find only the lower lying saddle points in order to identify the relevant transitions. The task of finding saddle points can be divided into two categories. In the more general category, only the initial state is known, and the task is to find all relevant saddle points on the energy ridge surrounding the energy minimum. In some cases, the final state of interest is also known, and the task can then be cast in terms of finding the minimum energy path (MEP) between the initial and final state minima. The MEP is a path for which the energy is at a minimum in directions orthogonal to the path. A maximum along the MEP corresponds to a first-order saddle point. The activation energy for the overall transition is then given by the highest rise in energy along the path. We will first discuss methods for finding the MEP when both initial and final states are given. Then, we discuss methods for the more challenging problem of finding relevant saddle points when only the initial state is specified.

## 3    Initial and Final States Specified, CI-NEB

A path between a given initial and final state can be discretized by creating a number of replicas of the system and arranging them in such a way as to trace out a curve between the states. We will refer to the discretization points as "images" of the system. An image labeled $i$ is specified by the coordinates of the atoms, $R_i$. The task is to first generate a reasonable initial path and then apply an optimization algorithm to iteratively move the images to the MEP. There can exist more than one MEP between the given minima, an issue that is addressed below. At this point, the task is just to find the MEP closest to the initial path.

The simplest way to generate an initial path is to make a linear interpolation in Cartesian coordinates between the two minima. This method is frequently used in materials simulations. It can, however, lead to unphysical configurations of the atoms. For example, two atoms may end up being very close to each other leading to a strong repulsive interaction. When the calculation of the energy and atomic forces is carried out using an electronic structure method, such strong overlap may slow down or even prevent the self-consistency calculation from converging. Furthermore, the initial path generated by a linear interpolation may be far from any MEP. A better approach is to generate the initial path by taking the pairwise distances between atoms into account. This can be done with the image dependent pair potential (IDPP) method (Smidstrup et al. 2014). There, the pairwise distances between neighboring atoms are interpolated linearly between the two minima and an initial path generated to match those distances as closely as possible. Since there are many more pairwise distances than atom coordinates, the matching can only be approximate and the initial path is found by minimizing the sum of squared deviations (Smidstrup et al. 2014). Another approach that avoids the problem of overlapping atoms is to generate the initial path by linear interpolation in internal coordinates (Goumans et al. 2009).

In order to find the MEP closest to the initial path, the images are moved iteratively in a direction obtained from the atomic forces, i.e., negative gradient of the energy with respect to atomic coordinates, $-\nabla E(R_i)$. But only the component perpendicular to the path should be used to modify the shape of the path. A force projection is, therefore, required based on an estimate of the local tangent to the path. While it seems natural to estimate the tangent at a given image, $i$, from the coordinates of atoms at the two adjacent images, $i - 1$ and $i + 1$, it turns out to be numerically more stable to use only the coordinates of the neighboring image that has the higher energy (Henkelman and Jónsson 2000). Letting the normalized tangent be denoted $\hat{\tau}_i$, the force acting on the shape of the path at image $i$ is given by

$$F_i^{\mathrm{p}}|_\perp = -\nabla E(R_i) + \nabla E(R_i) \cdot \hat{\tau}_i \, \hat{\tau}_i. \tag{3}$$

By displacing the images in the direction of $F_i^{\mathrm{p}}|_\perp$, the perpendicular component of the atomic forces will vanish, and the images are then on the MEP. At each iteration of the optimization, the atomic forces of all images need to be calculated. The evaluation of the energy and atomic forces is typically the most computationally intensive part of the calculation. However, the calculations can readily be performed simultaneously, using parallel or distributed computing.

It is also necessary to specify how the images are distributed along the path. If the distribution is not controlled, the images tend to slide down to the local energy minima. The distribution is commonly controlled using a restraint method where a harmonic spring acts between adjacent images in the direction parallel to the path:

$$F_i^{\mathrm{s}}|_\parallel = \left( k_i^s \, (R_{i+1} - R_i) - k_{i-1}^s \, (R_i - R_{i-1}) \right) \cdot \hat{\tau}_i \, \hat{\tau}_i. \tag{4}$$

The spring constants, $k_i^{\mathrm{s}}$, can be chosen to produce a desired distribution of the images, for example, with higher density in regions of higher energy (Henkelman et al. 2000a). Most often, though, the spring constants are chosen to have the same constant value resulting in an even distribution of images along the path. For optimal convergence rate, the magnitude of the spring constant should be chosen such that the magnitude of $F_i^{\mathrm{p}}|_\perp$ and $F_i^{\mathrm{s}}|_\parallel$ is comparable. The effective force acting on image $i$ in the path is then given by the sum of the two force components:

$$F_i = F_i^{\mathrm{p}}|_\perp + F_i^{\mathrm{s}}|_\parallel. \tag{5}$$

The force projections, referred to as "nudging," separate the distribution of the images along the path from the displacements of images affecting the shape of the path. The distribution of the images along the path can also be controlled using a constraint method based on an estimate of the total length of the path (Weiqing and Vanden-Eijnden 2002).

The most important part of the MEP is the highest energy point along the path, i.e., the highest first-order saddle point along the path connecting the initial and final

states. The activation energy of transitions between the two states can be estimated from the first-order saddle point using HTST. In order to obtain an accurate estimate of the saddle point, one of the images can be forced to "climb" upward along the path and converge to the highest saddle point. This image is referred to as the "climbing image." Generally, the highest energy image after a few normal iterations (or a certain user-defined threshold) is chosen to become the climbing image, $i = c$. The effective force acting on this image is (Henkelman et al. 2000a)

$$F_c = -\nabla E(R_c) + 2 \left( \nabla E(R_c) \cdot \hat{\tau}_c \right) \hat{\tau}_c \tag{6}$$

The assumption here is that the tangent estimate at the climbing image gives an accurate estimate of the direction of the unstable mode at the saddle point. It is used to transform the gradient of the energy around the first-order saddle point so as to mimic the gradient around a minimum. An ordinary minimization algorithm can then be used with the transformed gradient to converge the climbing image on the first-order saddle point. This method is referred to as the climbing image nudged elastic band (CI-NEB) method (Henkelman et al. 2000a).

## 3.1    Optimization of the Path

Various iterative algorithms can be used to zero the effective force and hence move the images to the MEP.

Note, however, that the objective function, the function that gives the effective force by differentiation, is not known explicitly because of the force projections, i.e., the nudging. The optimization algorithm should, therefore, be based only on the forces, not on the objective function itself. This, for example, makes the usual implementation of line search approaches not applicable and hence complicates the implementation of some of the optimization methods.

In early implementations of the method (Mills et al. 1995; Jónsson et al. 1998), a velocity projection method based on the velocity Verlet dynamics algorithm (Andersen 1980) was used. There, the velocity, $v$, is zeroed in directions orthogonal to the force if $v \cdot F > 0$ and in all directions if $v \cdot F < 0$. A parametrized extension of the algorithm has been devised (Bitzek et al. 2006). The velocity projection algorithm gives fast and reliable performance in the initial stages where the images may be located far from the MEP and the effective forces large (especially when linear interpolation is used to generate the initial path). However, the convergence close to the MEP can be faster by using more advanced approaches (Chu et al. 2003; Sheppard et al. 2008). The limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) (Nocedal 1980) algorithm has proved to be efficient (Sheppard et al. 2008). If the second derivatives of the energy are available and can be computed in each iteration without large effort (which however is generally not the case), a highly efficient optimization algorithm can be used (Bohner et al. 2013).

An example of a CI-NEB calculation on a modified two-dimensional Müller-Brown energy surface using the L-BFGS algorithm is shown in Fig. 1. The algorithm
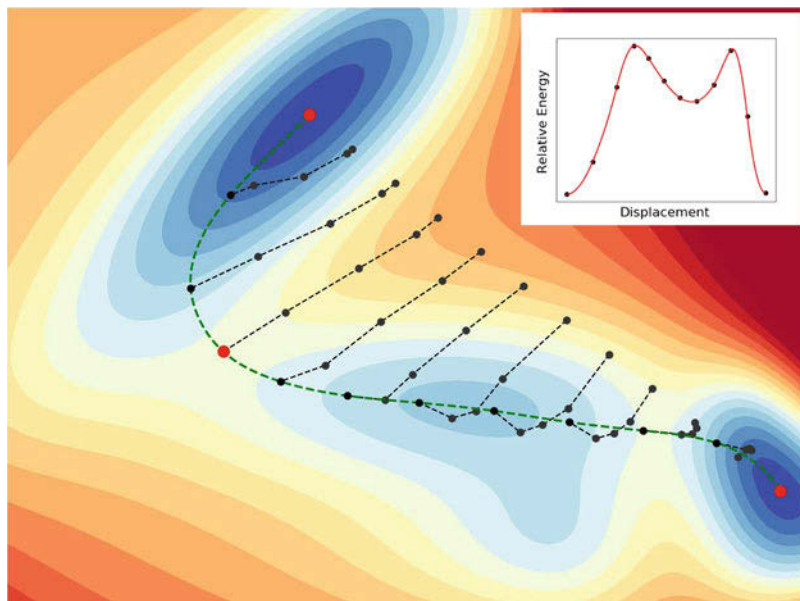
**Fig. 1** A climbing image nudged elastic band calculation of the minimum energy path of a model two-dimensional potential surface (a modified Müller-Brown surface (Ásgeirsson and Jónsson 2018)). In addition to the initial and final states, which are fixed (red dots), the path is represented by ten discretization points, referred to as images (black dots). The initial path is generated from a straight line interpolation (in more realistic systems, it is better to use the IDPP method). The position of images is shown after 0, 5, 10, 15, and 21 L-BFGS iterations. The green dashed line shows the minimum energy path. A red dot marks the first-order saddle point to which the climbing image converges. Inset: Energy along the minimum energy path and energy of the images after convergence has been reached

is used here right from the start. A more reliable strategy is to start out with the more stable and conservative velocity projection algorithm and then switch to the faster L-BFGS method once the effective force acting on each atom has dropped below some user-defined threshold, for example, $||F_i|| < 0.5\,\text{eV/Å}$.

A calculation may be considered tightly converged to the MEP when the norm of the force acting on each atom perpendicular to the path $||F_i^{\text{p}}|_\perp||$ is below $0.01\,\text{eV/Å}$ for all images. However, the MEP is mainly needed to identify where the highest energy point is and to estimate the tangent to the path at that point. The tight convergence is really only needed for the climbing image, and a looser threshold can be used for all other images, such as $||F_i^{\text{p}}|_\perp|| < 0.1\,\text{eV/Å}$.

When electronic structure methods, such as density functional theory (DFT) or ab initio quantum chemistry, are used in combination with a CI-NEB calculation, it is particularly important to reduce as much as possible the number of energy and atomic force evaluations. Recently, it has been shown that machine learning approaches can reduce the number of energy and force evaluations by an order of magnitude. Both neural networks (Peterson 2016) and Gaussian process regression

(GPR) (Koistinen et al. 2016, 2017) have been applied. There, an approximate energy surface is constructed with a machine learning model using all previous energy and force evaluations carried out by the electronic structure method. A CI-NEB calculation is carried out on the approximate energy surface and the electronic structure calculation then done for each image or only the image where the estimated uncertainty of the approximate energy surface is largest. The approximate energy surface is then refined using the new information and the CI-NEB calculation carried out again, etc. The GPR has the advantage of having a built-in error estimate which can be used to choose in an optimal way which image should be evaluated by the electronic structure method. The machine learning approach helps make optimal use of each electronic structure calculation, while traditional optimization methods, such as the velocity projection algorithm, only use the forces obtained at the current step and disregard all previous force evaluations. The L-BFGS keeps memory of the last $M$ steps (typically $M = 20$) to construct an approximate Hessian matrix, but the machine learning algorithm constructs an approximate surface of a more general shape. As a result, the machine learning approach can use all previous energy and force calculations and reach convergence with fewer electronic structure calculations than L-BFGS.

## 3.2    Interpolation and Strategy

In order to analyze and visualize the results of a CI-NEB calculation as well as to monitor the progress of the calculation, it is important to not just interpolate the energy of the images but also the derivative of the energy along the path, i.e., the component of the atomic forces parallel to the path. It is convenient to use a cubic polynomial for each interval between adjacent images (Henkelman and Jónsson 2000). An interpolation using the derivative may reveal the presence of an intermediate minimum along the path, while a simple interpolation of only the energy does not. A good strategy in such a case is to locate the energy minimum with a separate minimization calculation and then proceed with CI-NEB calculations of the two segments of the path separately.

In an analogous manner, the atom coordinates can be interpolated to add new images and hence improve the resolution in certain regions along the path (Henkelman and Jónsson 2000). This especially applies to regions of high curvature and rapid changes in the tangent from one image to another which can cause inaccuracies and even non-convergence of the calculation. As any other numerical method that relies on discretization, the CI-NEB method requires the number of images to be large enough. However, instead of doubling the number of images in such a situation, it may be sufficient to simply add an image (or a few images) in the problematic region along the path. Note that the spring constants associated with the new image need to be twice as strong such that the location of the adjacent images does not get disrupted.

When the CI-NEB method is applied to materials where periodic boundary conditions are applied to the simulation cell, the size and shape of the cell can be

made part of the MEP calculation, in addition to the atom coordinates (Sheppard et al. 2012). This is needed if the crystal structure is changing during the transition. When the method is applied to finite systems such as clusters or molecules, it is important to remove the overall translation and rotation of the system from the available degrees of freedom. The optimization of the path could otherwise lead to an artificial lengthening of the path involving translation and/or rotation in order to enable images to slide down from high-energy regions. This can reduce the resolution of the path in regions of high energy and increase the computational effort or even prevent the calculation from converging. A method based on quaternions has been formulated for the purpose of constraining the translation and rotation (Melander et al. 2015).

## 3.3 Application of CI-NEB and WKE Dynamics

An application of the CI-NEB method is shown in Fig. 2. It illustrates how the method can reveal a mechanism that is entirely different from the one implicitly assumed in the initial path. Also, after the MEP and relevant saddle point had been
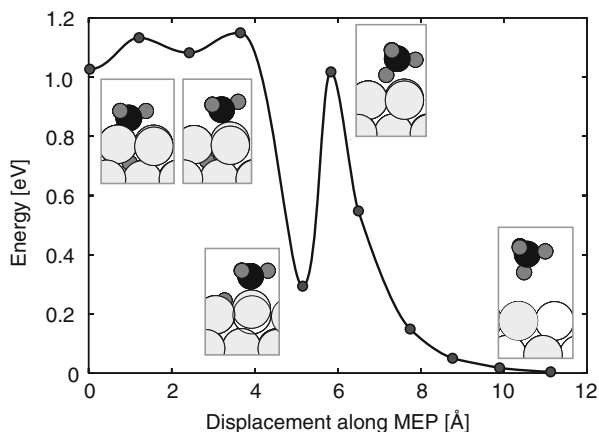


**Fig. 2** An application of the climbing image nudged elastic band method to a surface reaction where a $CH_4$ molecule is formed on a Ni(111) surface from a subsurface H atom and a $CH_3$ admolecule. While such a transition mechanism had been postulated to be efficient, the calculation shows that a corresponding minimum energy path does not exist on the DFT/PW91 energy surface. While the initial path is consistent with direct addition of the subsurface H atom with the $CH_3$ admolecule, the optimization of the path reveals that the $CH_3$ first hops to the side, the H atom emerges to the surface to a state that corresponds to a deep intermediate minimum, and, finally, the H adatom and the $CH_3$ combine in a regular surface reaction mechanism to form the $CH_4$, which leaves the surface. The calculation illustrates how the method can find at a complex minimum energy path involving several elementary reaction steps even when the initial path is generated with an incorrect mechanism in mind. (From Henkelman et al. 2006)

located, short time scale dynamics were initiated from the transition state to obtain various information about the atomic dynamics.

There had been some suggestions that subsurface H atoms could play an important role in the formation of $CH_4$ molecules from $CH_3$ molecules adsorbed on a catalyst surface. As can be seen from the inset showing the initial state, it seems reasonable that the H atom could add directly to the $CH_3$ as it moves up from the subsurface site. This mechanism appears to be more facile than the typical surface reaction mechanism since an H adatom on the surface cannot easily approach to the C atom in the $CH_3$ admolecule. A CI-NEB calculation for this reaction on a Ni(111) surface was carried out using DFT/PW91 to evaluate the energy and atomic forces (Henkelman et al. 2006). The calculation started from a linear interpolation assuming that the direct addition of the subsurface H atom to the C atom in $CH_3$ could take place. The initial path is simple and mainly involves the displacement of the H atom up from the subsurface site. Optimization of the path, however, revealed that an MEP corresponding to this direct reaction mechanism does not exist on the DFT/PW91 energy surface. Instead, the path converged on a complex multistep mechanism. First, the $CH_3$ hops to the side to a nearby bridge site, freeing up the surface site for the H atom. Then, the H atom hops to the surface. A deep minimum corresponds to this intermediate configuration. Finally, the H adatom and the $CH_3$ admolecule combine in the usual surface reaction step to form a $CH_4$ molecule that leaves the surface.

This is a good example of how a CI-NEB calculation can undo an incorrect preconceived notion of a reaction mechanism. While it can seem plausible that a subsurface H atom attaches directly to a $CH_3$ admolecule, the results obtained for the DFT/PW91 energy surface do not support this notion. The reason appears to be that the Ni atoms do not have catalytic activity where they are bonded to each other. The catalytic activity is confined to the side where the Ni atoms are exposed, i.e., to the surface.

Assuming the system is thermalized in the deep intermediate minimum corresponding to the H adatom and the $CH_3$ admolecule, the critical energy barrier for forming $CH_4$ is the one corresponding to the last saddle point. An interesting observation made from the CI-NEB calculation was a large displacement of the underlying Ni atom, by 0.25 Å. While a similar observation had been made in calculations of $CH_4$ formation on Ir(111) (Henkelman and Jónsson 2001), most surface reaction calculations were at that time carried out with frozen surface atoms.

This raises interesting questions about the atomic dynamics of the transition. For example, how is the 0.85 eV excess energy gained by the system as it advances past the transition state to a gas phase $CH_4$ molecule and clean Ni surface partitioned between the Ni atoms and the $CH_4$ molecule? Also, how is the energy of the $CH_4$ molecule partitioned among its vibrational modes? For the reverse reaction, dissociative sticking of $CH_4$, this reveals to what extent translation and the various vibrational modes of the molecule can enhance the sticking probability. To answer these questions, the second step of the WKE approximation was carried out, using the HTST approximation of the transition state. Only a minimal sampling of trajectories was carried out, starting with small displacements from the saddle

point. The calculations showed that 15% of the excess energy go into Ni atom vibrations, while 79% go into translation, 22% into vibrations, and 4% into rotation of the $CH_4$ molecule. Surprisingly, most of the vibrational energy goes into the stretching modes not the lower frequency deformation modes. This result indicates that excitation of the stretching modes is more likely to enhance sticking than excitation of the deformation modes. The translation is, however, by far most effective in enhancing sticking.

Another interesting dynamics issue that could be addressed in the second step of the WKE procedure is whether the high kinetic energy of the H adatom as it emerges at the surface could enhance the reaction with the $CH_3$. To address this question, dynamical trajectories started at the transition state for the subsurface to surface hop were calculated, but none of the trajectories led to a combination of the H atom with the $CH_3$. The conclusion is that subsurface H atoms do not play an important role in hydrogenation of $CH_3$ admolecules under the conditions simulated.

## 3.4    Variants of the Method

There are several different variants of the method for finding an MEP when both initial and final states are specified, and a few of them will be mentioned here briefly.

When the shape of the MEP has become clear enough for the approximate location of the highest energy saddle point to be located approximately, computational time can be saved by confining the calculations to the region in the vicinity of the saddle point and halting calculations of images in other regions. There have been several different implementations of this approach. In its simplest form, two images on opposite sides of the highest energy image of the partly converged path are chosen to be the fixed endpoints instead of the local minima, but a better approach is to make the new endpoints follow selected equipotential contours to the MEP (Zhu et al. 2007; Zhang et al. 2016). Another approach is to use two climbing images located on opposite sides of the highest energy image so as to bracket the saddle point (Zarkevich and Johnson 2015).

Alternatively, a single-ended saddle point search, discussed below, can be launched from the highest energy image on the partially converged path (Henkelman et al. 2000b).

When dealing with complex energy surfaces with multiple local minima and curved MEPs, the stability of the calculation can be improved by including part of the component of the spring force perpendicular to the path (Jónsson et al. 1998) or by using the so-called double nudging (Trygubenko and Wales 2004; Sheppard et al. 2008). Also, long paths and large number of images can cause the distance between images to become uneven. In such cases, a combination of the restraint and constraint approaches for distributing the images along the path can be used (Maras et al. 2016, 2017). Alternatively, it is good to divide the path calculation and focus on each segment separately whenever there is a sign of an intermediate minimum, as discussed above. An automatic algorithm for adding new images to the discrete representation of the path has been developed (Kolsbjerg et al. 2016).

# 4    Only Initial State Specified, MMF

It can be hard to predict the mechanism and even the possible final states that can be formed in a thermally activated transition. Ideally, a computer calculation should be able to reveal such information given only the initial state of the system. This is a harder problem than the one discussed above where the final state is specified as well as the initial state. One option is to raise the temperature of the system and identify new states that get visited, followed by an NEB calculation to find the minimum energy path from the initial state (Sørensen et al. 1996). But the increase in temperature will place larger emphasis on states with high entropy and the transitions relevant at the low temperature of interest may be hard to find. What is needed is a method for climbing up the energy surface from the initial state to converge onto a first-order saddle point. Such searches need to be carried out several times in order to identify all relevant, i.e., low-lying saddle points on the energy ridge surrounding the initial state minimum. Armed with such a method, the long time scale evolution of a materials system can be simulated, as discussed in chapter "The Chapter on AKMC Methods".

A first-order saddle point differs from a local minimum in that the Hessian matrix has one negative eigenvalue, corresponding to the unstable mode. The atomic forces near a first-order saddle point can be made to mimic atomic forces near a local minimum by reversing the component of the force in the direction of the unstable mode. More generally, let $\hat{\lambda}$ be a unit vector along the eigenvector corresponding to the lowest eigenvalue of the Hessian, the minimum mode (irrespective of the sign of the lowest eigenvalue). If the system is made to follow an effective force given by

$$F^{\text{eff}}(R) = -\nabla E(R) + 2\left(\nabla E(R) \cdot \hat{\lambda}(R)\right)\hat{\lambda}(R) \tag{7}$$

then convergence will be reached at a first-order saddle point. This is analogous to the force on the climbing image in the CI-NEB algorithm except that $\hat{\lambda}$ is used here instead of the path tangent, which is not known in this case. If the second derivatives of the energy can be evaluated easily, then the Hessian matrix can be constructed and its eigenvalues and eigenvectors computed quite readily, although the computational effort can be significant for large systems, as it scales with the third power of the number of degrees of freedom.

In most cases, however, the second derivatives of the energy cannot be obtained easily, and the challenge is to evaluate $F^{\text{eff}}$ using only the first derivatives, i.e., the atomic forces, and without even evaluating the Hessian matrix. There are several ways to do this. One is to construct a dimer, i.e., two replicas of the system held at a small, fixed distance apart. The lowest energy orientation of the dimer is the direction of the minimum mode, $\hat{\lambda}$, and it can be found by iterative minimization of the energy using only the first derivatives (Henkelman and Jónsson 1999). An analogous method was formulated at a similar time (Munro and Wales 1999). Another way to find the minimum mode is to use the iterative Lanczos or,

even better, the Davidson method (Malek and Mousseau 2000; Olsen et al. 2004; Gutiérrez et al. 2016).

Since the system is displaced uphill in energy along the minimum mode, the method is often referred to as minimum mode following (MMF). The first step is to displace the atoms slightly from the initial state minimum and drive the system in some way into the region where the lowest eigenvalue of the Hessian is negative. From then on, the system is displaced in the direction of $F^{\text{eff}}$ until the magnitude of the force drops below a given tolerance. The minimum mode will often change only slightly between iterations. Therefore, it is possible to save computational effort by only evaluating the minimum mode every few iterations (Gutiérrez et al. 2016). The MMF method is discussed in more detail in chapter "The Chapter on AKMC Methods".

Figure 3 shows MMF calculations for two-dimensional test problems. There, a random displacement from the initial state minimum is first applied, and then it is advanced in the direction of the minimum mode until the lowest eigenvalue becomes
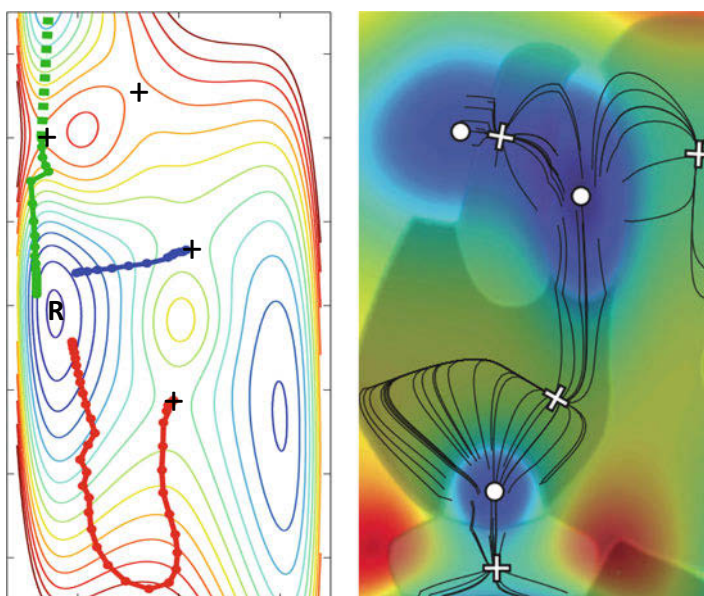


**Fig. 3** An illustration of the minimum mode following method for finding first-order saddle points given only the initial state minimum (blue). Saddle points are marked with + signs. Left: Three saddle point searches started from different, random displacements from the minimum. In one of the searches (green), a kink on the path is evident when the lowest eigenvalue becomes negative and the perpendicular relaxation is turned on. After locating the saddle point, the system is displaced in the direction of the unstable mode, followed by minimization to locate the final state minimum. Right: A larger number of saddle point searches started from three different initial states. The region around each saddle point where at least one eigenvalue is negative is illustrated with a lighter color. In some cases, several different search paths trace the same final approach to the saddle point

negative. From that point on, the system is advanced in the direction of $F^{\text{eff}}$, and relaxation perpendicular to the minimum mode is included. This can result in a kink in the path. The figure also shows that several paths can approach a saddle point in a similar way even when they start out being quite different.

An application to a materials problem is shown in Fig. 4 where various different migration mechanisms of a kink on a dislocation in silicon crystal are identified (Pedersen et al. 2009a). The calculation made use of the EDIP empirical potential function (Justo et al. 1998). Three different mechanisms were found from multiple saddle point searches. The optimal mechanism, the one corresponding to lowest activation energy, involves an intermediate state where three Si atoms have fivefold coordination. Two of the mechanisms, including the optimal one, involve an intermediate state.
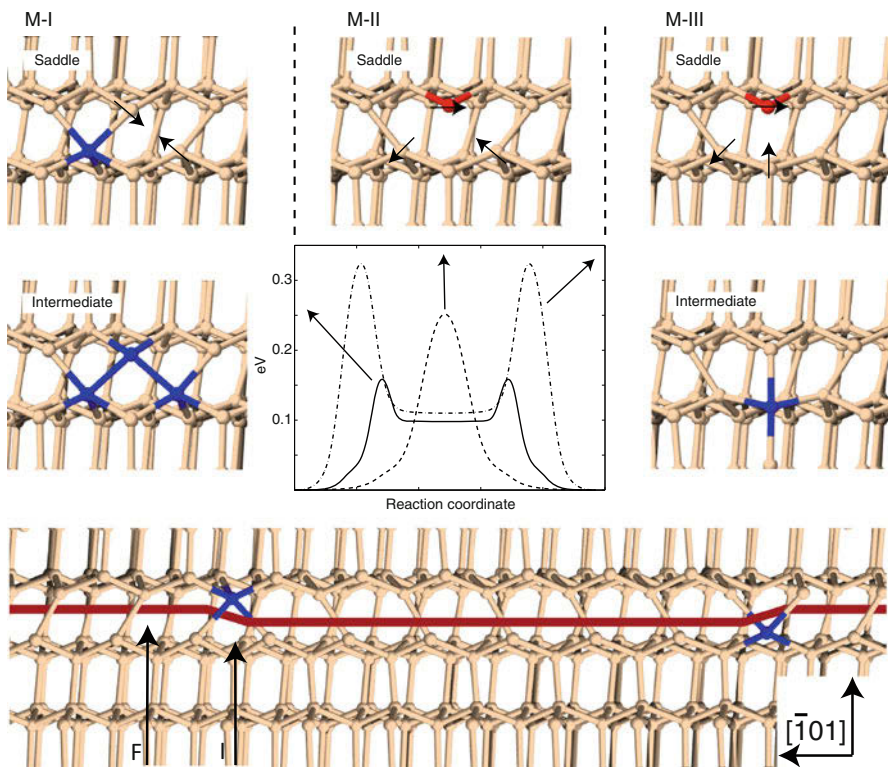


**Fig. 4** Application of the minimum mode following method to find the mechanism of dislocation kink migration in crystalline silicon, based on the EDIP empirical potential function. Fivefold coordinated atoms are shown in blue and threefold coordinated atoms in red. The three lowest energy saddle points are shown (top row), two of them lead to an intermediate state (middle row). The initial state with two kinks separated by a large distance is shown (bottom), with a red line indicating the dislocation. (From Pedersen et al. 2009a)

Minimum mode saddle point searches have been used in various studies, for example, of the binding and diffusion of $H_2O$ molecules on the surface of proton-disordered ice (Pedersen et al. 2014, 2015) and structure and diffusion in metal grain boundaries (Pedersen et al. 2009b; Pedersen and Jónsson 2009). Several other applications are discussed in chapter "AKMC Methods".

## 5 Additional Characterization of the Energy Surface

In this section, a few additional saddle point search methods that can be used to further characterize the energy surface of a system are discussed.

### 5.1 Search for Optimal MEPs

So far, the discussion of the MEP calculations has focused on finding only the MEP closest to the initial path. Sometimes the energy surface is simple enough, and the initial and final states are similar enough that only a single MEP connects the two. But, more generally, there can be two or more MEPs for a given pair of endpoints. One example is the $CH_4$ formation shown in Fig. 3. The $CH_3$ admolecule could jump to the left instead of jumping to the right in the first step to make room for the H atom on the surface. Also, the subsurface H atom could jump to an adjacent subsurface site before going to the surface, analogous to an MEP found for $H_2$ formation (Henkelman et al. 2006). These various paths would all have the same critical step, namely, the combination of an H adatom and a $CH_3$ admolecule on the surface, so they would, for practical purposes, be equivalent.

For other systems, it may be important to sample the MEPs to find the one that corresponds to lowest overall activation energy. One example of such a system is the nucleation of dislocation on a strained Ge overlayer (Maras et al. 2017). Since the atomic model of the dislocation involves a large number of atoms, there are many possible mechanisms, and each one involves a large number of intermediate minima. A global optimization strategy based on an evolutionary algorithm has been applied to this problem (Maras et al. 2017). There, segments of different paths are mixed and matched and a catalog maintained of all the intermediate minima found so far. More work needs to be done to develop optimal strategies for such complex systems.

### 5.2 Energy Ridge Tracking

The accuracy of the HTST approximation rests on the assumption that second-order saddle points are significantly higher in energy, on the scale of $k_BT$, than the first-order saddle points. In order to check that, a calculation of a path lying along the energy ridge between two first-order saddle points can be carried out. The method, ridge tracking nudged elastic band (RT-NEB), is an extension of the

CI-NEB method where each discretization point is a dimer of images (Maronsson et al. 2012). It combines, in a sense, the technique discussed above for finding an MEP and the dimer method for finding a saddle point given only the initial state. Just as the CI-NEB method can reveal the presence of intermediate minima that were not known beforehand, the RT-NEB method can reveal the presence of first-order saddle points in between known saddle points. It can, therefore, help complete the table of possible transitions a system can undergo (Maronsson et al. 2012).

## 5.3    Saddle Points for Quantum Tunneling

At low enough temperature, quantum mechanical tunneling becomes the dominant transition mechanism rather than hops over the energy barrier. An estimate of the crossover temperature can be obtained from the MEP (Gillan 1987). The larger the curvature of the MEP and the lower the effective mass along the unstable mode at the saddle point, the higher the crossover temperature.

The quantum statistical mechanics of the system can be obtained conveniently from thermal path integrals (Feynman and Hibbs 1965). Within that formalism, the classical energy surface gets replaced by an energy surface for a ring polymer where the beats are replicas of the system connected with temperature-dependent springs. This gives an effective, quantum mechanical energy surface that depends on temperature. The higher the temperature and the larger the mass of the atoms, the stiffer the springs and more classical the system becomes (Feynman and Hibbs 1965). A quantum mechanical extension of transition state theory can be formulated in terms of the rate of transitions of the ring polymer from the initial state to the final state (Mills et al. 1997, 1998; Richardson and Althorpe 2009; Hele and Althorpe 2013). Within a harmonic approximation, the optimal tunneling path is a first-order saddle point on the quantum mechanical energy surface (Benderskii et al. 1994; Mills et al. 1998), and the rate of thermally activated tunneling can be estimated from its properties (Benderskii et al. 1994; Richardson 2016). The optimal tunneling path is often referred to as an "instanton."

Saddle point search methods, such as the MMF method, can be used to find instantons (Andersson et al. 2009; Rommel and Kästner 2011). However, a more efficient approach is to use a path optimization method that minimizes a line integral along the path. This line integral nudged elastic band (LI-NEB) method can be used to find the optimal tunneling path at a given energy based on the semiclassical JWKB approximation. There, the images can be distributed evenly. Then, after the path has been found, a larger number of images need to be distributed unevenly along the path to represent an instanton, the optimal Feynman path for tunneling at a given temperature (Einarsdóttir et al. 2012; Ásgeirsson and Jónsson 2018). The beads of the ring polymer have a higher density in the lower-energy regions, making the direct search for the instanton using the MMF method less efficient than the LI-NEB.

Examples of tunneling paths at two different values of the temperature are shown in Fig. 5 for the same two-dimensional energy surface as in Fig. 1. The calculation
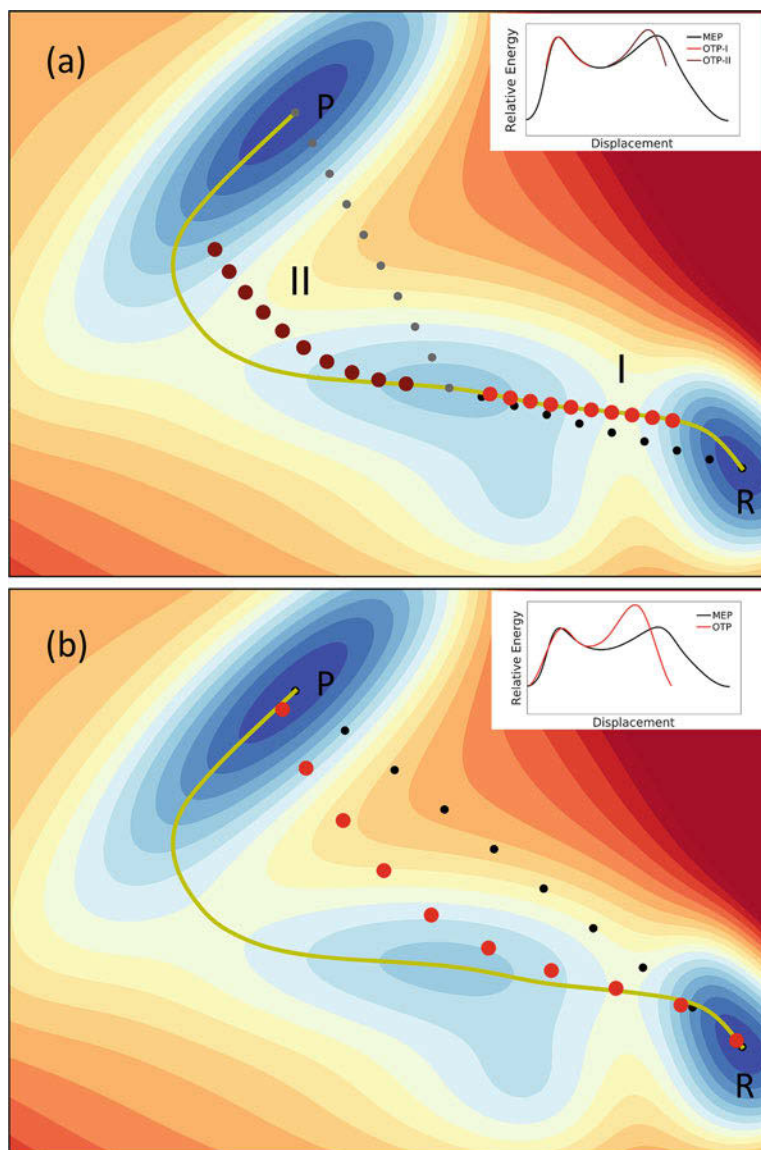
**Fig. 5** Optimal tunneling paths found using the line integral nudged elastic band method. The energy surface is the same as in Fig. 1. (**a**) Tunneling occurs in two steps, first into the intermediate state and then the final state. The initial, straight line interpolation path is shown with black and gray dots, while the optimal tunneling paths are shown with red and brown dots. The corner cutting of the tunneling path is particularly clear at saddle point II where the MEP has a larger curvature. (**b**) Tunneling occurs in one step between states R and P, without visiting the intermediate state. At low enough temperature, this becomes the dominant transition mechanism. The initial, straight line interpolation path is shown with black dots, while red dots show the optimal tunneling path. Insets: Energy along the minimum energy path and along the optimal tunneling paths. (From Ásgeirsson and Jónsson 2018)

of the first-order saddle points on the quantum mechanical surface, the instantons, serves as further explorations of the energy surface. In regions where the MEP is curved, the tunneling path cuts corners. At the higher temperature, the tunneling takes place in two steps, first from the initial state, R, to the intermediate state and then from there to the final state, P. At the lower temperature, the tunneling takes the system directly between states R and P without entering the intermediate state. Results of the LI-NEB calculation are shown where the images are evenly distributed along the optimal tunneling path.

## 6    Saddle Points for Magnetic Transitions

Transitions from one magnetic state of a material to another can also be slow on the time scale of the vibrations of magnetic moments (Braun 2012). The same considerations about time scale differences and rare events discussed above in the context of atomic rearrangements can apply to such transitions. As a result, the characterization of the energy surface as a function of the magnetic degrees of freedom using saddle point searches is also useful when studying magnetic systems. It turns out that a semiclassical, adiabatic approximation can usually be made where the angles specifying the orientation of the magnetic momentum vectors are treated as slow variables while the length of the magnetic momentum vectors is a fast variable, determined by the electronic structure (Antropov et al. 1996). A full semiclassical specification of the configuration of an atomic scale system should, therefore, include the location of the atomic nuclei as well as two angles, for example, the polar and azimuthal angles ($\theta$ and $\phi$) for each atom. A transition can involve simultaneous rearrangement of atoms and reorientation of magnetic moments. Within a harmonic approximation, the mechanism of such a transition can be characterized by a saddle point on this higher-dimensional energy surface.

A brief account will be given here of tools for navigating on magnetic energy surfaces, analogous to the tools presented above for atomic rearrangements. For simplicity, the atom coordinates will be assumed to be fixed and only the orientation of the magnetic moments treated as variables. The configuration space of N magnetic moments is a direct product of N two-dimensional spheres. The fact that the configuration space is curved brings up special considerations for the saddle point search methods. Again, the activation energy for thermally activated transitions is given by the energy difference between a first-order saddle point and the initial state minimum. But the equation of motion for magnetic moments, the Landau-Lifshitz equation, gives a different pre-exponential factor in the HTST approximation for the rate constant (Bessarab et al. 2012, 2013).

### 6.1    Geodesic CI-NEB Method

The distance between two points in the configuration space of magnetic moments is given by the length of the geodesic path connecting the two points. This should be

taken into account in the CI-NEB calculations of minimum energy paths between two local minima on the magnetic energy surface. Also, the nudging needs to include a projection onto the tangent space of the N two-dimensional spheres. This variant of the method for finding MEPs for magnetic transitions is referred to as the geodesic nudged elastic band (GNEB) method (Bessarab et al. 2015).

An example of a calculation of an MEP for a two-dimensional magnetic system is shown in Fig. 6. The Hamiltonian here is of an extended Heisenberg form including the Dzyaloshinsky-Moriya interaction  (Uzdin et al. 2018). The system initially
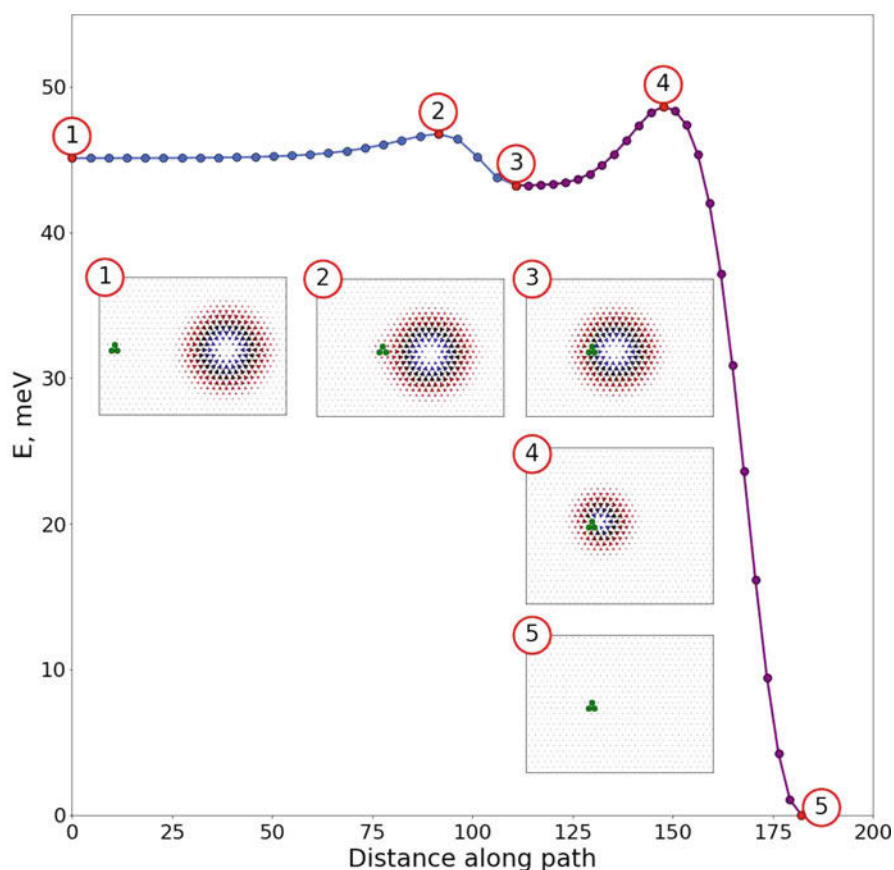


**Fig. 6** Geodesic climbing image nudged elastic band method calculation of a minimum energy path for a magnetic transition where a skyrmion first attaches to a three-atom nonmagnetic impurity and then collapses. Red (blue) arrows show spins with z-component pointing up (down). Initially, the skyrmion and the nonmagnetic impurity are separated by a large distance, (1). By overcoming a small energy barrier, (2), a more stable state is reached where the skyrmion incorporates the impurity in a region where its magnetic moments lie in the plane (black arrows), (3). The skyrmion then collapses by overcoming a larger energy barrier (4) resulting in the ground-state, ferromagnetic phase (5). (From Uzdin et al. 2018)

contains a localized non-collinear structure called a skyrmion. There is great interest in such states as they represent a pseudo-particle that can be quite stable and could be used for future data storage and even data manipulation. The MEP shows the skyrmion approaching a triatomic, nonmagnetic impurity. By overcoming a small energy barrier, the impurity becomes incorporated in the skyrmion, in a region where the magnetic moments lie in the plane of the system. This lowers the energy of the system slightly. While a skyrmion in a continuous field is rigorously stable due to topological protection, its collapse can occur with a finite energy barrier on a discrete lattice (Bessarab et al. 2015). Here, the collapse occurs in the presence of the impurity. The mechanism involves shrinkage of the skyrmion and finally rotation of the core. For this set of parameters, the uniform ferromagnetic phase is lower in energy.

## 6.2    Mode Following in Curved Space

Since transitions involving reorientations of magnetic moments can be complex, many different types of transitions can be possible; a saddle point search method where only the initial state is specified is useful, just as for atomic rearrangements. The curvature of the configuration space needs to be taken into account when constructing the Hessian and calculating the eigenvalues and eigenvectors. An extension of mode following has recently been formulated for magnetic systems and applied to skyrmions (Müller et al. 2018). In addition to collapse and escape through a boundary, the method identified a transition where the skyrmion divides and forms two equivalent skyrmions. The activation energy for this duplication process was for some parameter values similar to that of collapse and escape. A dynamics simulation based on the Landau-Lifshitz equation showed that a time-dependent external field could induce the duplication process (Müller et al. 2018).

   More work is needed to complete the implementation of the saddle point searches for systems of atoms with magnetic moments. First of all, an efficient method for constraining spin orientations in nonstationary configurations in electronic structure calculations needs to be developed. Then, a formulation of these various saddle point search methods for energy surfaces that are functions of both atomic coordinates and magnetic momentum orientations needs to be developed and implemented.

## 7    Conclusion

This chapter has summarized briefly various frequently used methods for finding saddle points. Such calculations can be used to reveal new states of the system by identifying local energy minima, not just saddle points. The height of the saddle point and the curvature of the energy surface around it as well as around the minimum can, within HTST, be used to estimate the rate of transitions between states. The simplest and most stable way to find saddle points is by finding MEPs

and identifying maxima along the path. This is now routinely done in the context of electronic structure calculations, DFT or ab initio. The MMF method where only the initial state is specified is more challenging and computationally demanding, and most calculations so far have been carried out with empirical potential functions. This will likely change in the future when more computational cores become available since the calculations involve several independent saddle point searches that can easily be carried out simultaneously via parallel or distributed computing (Chill et al. 2014a). Machine learning such as Gaussian process regression will also likely reduce the number of electronic structure calculations, as has already been demonstrated in CI-NEB calculations. Saddle point searches may even become an important component of global optimization algorithms (Pedersen et al. 2012; Plasencia et al. 2014). It is important to keep improving the algorithms to reduce the computational effort. A web page that collects algorithms and performance numbers for saddle point searches has been set up and is a valuable resource in this field (Chill et al. 2014b).

# References

Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 72(4):2384–2393

Andersson S, Nyman G, Arnaldsson A, Manthe U, Jónsson H (2009) Comparison of quantum dynamics and quantum transition state theory estimates of the H + CH$_4$ reaction rate. J Phys Chem A 113:4468

Antropov VP, Katsnelson MI, Harmon BN, van Schilfgaarde M, Kusnezov D (1996) Spin dynamics in magnets: equation of motion and finite temperature effects. Phys Rev B 54:1019

Ásgeirsson V, Jónsson H (2018) Efficient evaluation of atom tunneling combined with electronic structure calculations. J Chem Phys 148:102334

Benderskii VA, Makarov DE, Wight CA (1994) Chemical dynamics at low temperatures. Adv Chem Phys 88:1

Bessarab PF, Uzdin VM, Jónsson H (2012) Harmonic transition state theory of thermal spin transitions. Phys Rev B 85:184409

Bessarab PF, Uzdin VM, Jónsson H (2013) Potential energy surfaces and rates of spin transitions. Zeitschrift für Physikalische Chemie 227:1543

Bessarab PF, Uzdin VM, Jónsson H (2015) Method for finding mechanism and activation energy of magnetic transitions, applied to skyrmion and antivortex annihilation. Comput Phys Commun 196:335

Bitzek E, Koskinen P, Gʹahler F, Moseler M, Gumbasch P (2006) Structural relaxation made simple. Phys Rev Lett 97(17):170201

Bligaard T, Jónsson H (2005) Optimization of hyperplanar transition states: application to 2D test problems. Comput Phys Commun 169:284

Bohner MU, Meisner J, Kastner J (2013) A quadratically-converging nudged elastic band optimizer. J Chem Theory Comput 9(8):3498–3504

Braun H-B (2012) Topological effects in nanomagnetism: from superparamagnetism to chiral quantum solitons. Adv Phys 61(1):1–116

Chill ST, Welborn M, Terrell R, Zhang L, Berthet J-C, Pedersen A, Jónsson H, Henkelman G (2014a) EON: software for long time simulations of atomic scale systems. Model Simul Mater Sci Eng 22:055002

Chill ST, Stevenson J, Ruehle V, Cheng S, Xiao P, Farrel JD, Wales DJ, Henkelman G (2014b) Benchmarks for characterization of minima, transition states, and pathways in atomic, molecular, and condensed matter systems. J Chem Theory Comput 10:5476–5482

Chu J-W, Trout BL, Brooks BR (2003) A super-linear minimization scheme for the nudged elastic band method. J Chem Phys 119(24):12708–12717

Ciccotti G, Ferrario M, Laria D, Kapral R (1995) Simulation of classical and quantum activated processes in the condensed phase. In: Reatto L, Manghi F (ed) Progress in computational physics of matter: methods, software and applications. World Scientific, Singapore, p 150

Einarsdóttir DM, Arnaldsson A, Óskarsson F, Jónsson H (2012) Path optimization with application to tunneling. Lect Notes Comput Sci 7134:45

Eyring H (1935) The activated complex in chemical reactions. J Chem Phys 3:107

Feynman RP, Hibbs AR (1965) Quantum mechanics and path integrals. McGraw-Hill, New York

Gillan MJ (1987) Quantum-classical crossover of the transition rate in the damped double well. Phys C Solid State Phys 20(24):3621–3641

Goumans TPM, Catlow CRA, Brown WA, Kästner J, Sherwood P (2009) An embedded cluster study of the formation of water on interstellar dust grains. Phys Chem Chem Phys 11(26):5431–5436

Gutiérrez MP, Argáez C, Jónsson H (2017) Improved minimum mode following method for finding first order saddle points. J Chem Theory Comput 13(1):125–134

Hele TJH, Althorpe SC (2013) Derivation of a true ($t \rightarrow 0+$) quantum transition-state theory. I. Uniqueness and equivalence to ring-polymer molecular dynamics transition-state-theory. J Chem Phys 138:084108

Henkelman G, Jónsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111:7010

Henkelman G, Jónsson H (2000) Improved tangent estimate in the NEB method for finding minimum energy paths and saddle points. J Chem Phys 113(22):9978–9985 [Note: There is a typographical error in the Appendix, $2V_{i+1} - V_i$ should be $-2(V_{i+1} - V_i)$]

Henkelman G, Jónsson H (2001) Theoretical calculations of dissociative adsorption of methane on an Ir(111) surface. Phys Rev Lett 86:664

Henkelman G, Uberuaga BP, Jónsson H (2000a) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904

Henkelman G, Jóhannesson G, Jónsson H (2000b) Theoretical methods in condensed phase chemistry, methods for finding saddle points and minimum energy paths. In: Schwartz SD (ed) Progress in theoretical chemistry and physics. Kluwer Academic Publishers, Dordrecht, pp 269–302

Henkelman G, Arnaldsson A, Jónsson H (2006) Theoretical calculations of $CH_4$ and $H_2$ associative desorption from Ni(111): could subsurface hydrogen play an important role? J Chem Phys 124:044706

Jóhannesson GH, Jónsson H (2001) Optimization of hyperplanar transition states. J Chem Phys 115:9644

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Berne BJ (ed) Classical and quantum dynamics in condensed phase simulations. World Scientific, Singapore, pp 385–404

Justo JF, Bazant MZ, Kaxiras E, Bulatov VV, Yip S (1998) Interatomic potential for silicon defects and disordered phases. Phys Rev B 58:2539

Keck JC (1967) Variational theory of reaction rates. J Chem Phys 13:85

Koistinen O-P, Maras E, Vehtari A, Jónsson H (2016) Minimum energy path calculations with Gaussian process regression. Nanosyst Phys Chem Math 7:925

Koistinen O-P, Dabgjartsdóttir F, Ásgeirsson V, Vehtari A, Jónsson H (2017) Nudged elastic band calculations accelerated with gaussian process regression. J Chem Phys 147(15):152720

Kolsbjerg EL, Groves MN, Hammer B (2016) An automated nudged elastic band. J Chem Phys 145(9):094107

Malek R, Mousseau N (2000) Dynamics of Lennard-Jones clusters: a characterization of the activation-relaxation technique. Phys Rev E 62(6):7723–7728

Maras E, Trushin O, Stukowski A, Ala-Nissila T, Jónsson H (2016) Global transition path search for dislocation formation in Ge on Si(001). Comput Phys Commun 205:13

Maras E, Pizzagalli L, Ala-Nissila T, Jónsson H (2017) Atomic scale formation mechanism of edge dislocation relieving lattice strain in a GeSi overlayer on Si(001). Sci Rep 7:11966

Maronsson JB, Jónsson H, Vegge T (2012) A method for finding the ridge between saddle points applied to rare event rate estimates. Phys Chem Chem Phys 14:2884

Melander M, Laasonen K, Jónsson H (2015) Removing external degrees of freedom from transition-state search methods using quaternions. J Chem Theory Comput 11(3):1055–1062

Mills G, Jónsson H, Schenter GK (1995) Reversible work based transition state theory: application to $H_2$ dissociative adsorption. Surf Sci 324:305–337

Mills G, Schenter GK, Makarov DE, Jónsson H (1997) Generalized path integral based quantum transition state theory. Chem Phys Lett 278:91

Mills G, Schenter GK, Makarov DE, Jónsson H (1998) RAW quantum transition state theory. In: Berne BJ, Ciccotti G, Coker DF (eds) Classical and quantum dynamics in condensed phase simulations. World Scientific, Singapore, pp 405–421

Müller GP, Bessarab PF, Vlasov FM, Lux F, Kiselev NS, Blügel S, Uzdin VM, Jónsson H (2018) Duplication, collapse and escape of magnetic skyrmions revealed using a systematic saddle point search method. Phys. Rev. Lett. 121:197202

Munro LJ, Wales DJ (1999) Defect migration in crystalline silicon. Phys Rev B 59:3969

Nocedal J (1980) Updating quasi-Newton matrices with limited storage. Math Comput 35(151):773–782

Olsen RA, Kroes G-J, Henkelman G, Arnaldsson A, Jónsson H (2004) Comparison of methods for finding saddle points without knowledge of the final states. J Chem Phys 121(20):9776–9792

Pedersen A, Jónsson H (2009) Simulations of hydrogen diffusion at grain boundaries in aluminum. Acta Materialia 57:4036

Pedersen A, Pizzagalli L, Jónsson H (2009a) Finding mechanism of transitions in complex systems: formation and migration of dislocation kinks in a silicon crystal. J Phys Condens Matter 21:084210

Pedersen A, Henkelman G, Schioetz J, Jónsson H (2009b) Long timescale simulation of a grain boundary in copper. New J Phys 11:073034

Pedersen A, Berthet J-C, Jónsson H (2012) Simulated annealing with coarse graining and distributed computing. Lect Notes Comput Sci 7134:34

Pedersen A, Wikfeldt KT, Karssemeijer LJ, Cuppen HM, Jónsson H (2014) Molecular reordering processes on ice (0001) surfaces from long timescale simulations. J Chem Phys 141:234706

Pedersen A, Karssemeijer LJ, Cuppen HM, Jónsson H (2015) Long-timescale simulations of $H_2O$ admolecule diffusion on Ice Ih(0001) surfaces. J Phys Chem C 119:16528

Peters B (2017) Reaction rate theory and rare events. Elsevier Science & Technology, Amsterdam

Peterson AA (2016) Acceleration of saddle-point searches with machine learning. J Chem Phys 145(7):074106

Plasencia M, Pedersen A, Arnaldsson A, Berthet J-C, Jónsson H (2014) Geothermal model calibration using a global minimization algorithm based on finding saddle points as well as minima of the objective function. Comput Geosci 65:110

Richardson JO (2016) J Chem Phys 144:114106

Richardson JO, Althorpe SC (2009) Ring-polymer molecular dynamics rate-theory in the deep-tunneling regime: connection with semiclassical instanton theory. J Chem Phys 131:214106

Rommel JB, Kästner J (2011) Adaptive integration grids in instanton theory improve the numerical accuracy at low temperature. J Chem Phys 134:184107

Sheppard D, Terrell R, Henkelman G (2008) Optimization methods for finding minimum energy paths. J Chem Phys 128(13):134106

Sheppard D, Xiao P, Chemelewski W, Johnson DD, Henkelman G (2012) A generalized solid-state nudged elastic band method. J Chem Phys 136(7):074103

Smidstrup S, Pedersen A, Stokbro K, Jónsson H (2014) Improved initial guess for minimum energy path calculations. J Chem Phys 140(21):214106

Sørensen MR, Jacobsen KW, Jónsson H (1996) Thermal diffusion processes in metal tip-surface interactions: contact formation and adatom mobility. Phys Rev Letters 77(25):5067–5070

Trygubenko SA, Wales DJ (2004) A doubly nudged elastic band method for finding transition states. J Chem Theory Comput 120(5):2082–2094

Uzdin VM, Potkina MN, Lobanov IS, Bessarab PF, Jónsson H (2018) Energy surface and lifetime of magnetic skyrmions. J Magn Magn Mater 459:236–240

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3:121

Voter AF, Doll JD (1985) Dynamical corrections to transition state theory for multistate systems: surface self-diffusion in the rare-event regime. J Chem Phys 82(1):80–92

Wigner E (1938) The transition state method. Trans Faraday Soc 34:29

Weinan E, Weiqing R, Vanden-Eijnden E (2002) String method for the study of rare events. Phys Rev B 66(4):052301

Zarkevich NA, Johnson DD (2015) Nudged-elastic band method with two climbing images: finding transition states in complex energy landscapes. J Chem Phys 142(2):024106

Zhang J, Zhang H, Ye H, Zheng Y (2016) Free-end adaptive nudged elastic band method for locating transition states in minimum energy path calculation. J Chem Phys 145(9):094104

Zhu T, Li J, Samanta A, Kim HG, Suresh S (2007) Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. Proc Natl Acad Sci USA 104(9):3031–3036

# Off-Lattice Kinetic Monte Carlo Methods

# 34

## Mickaël Trochet, Normand Mousseau, Laurent Karim Béland, and Graeme Henkelman

## Contents

M. Trochet (✉)

Departement de physique and Regroupement québécois sur les matériaux de pointe, Université de Montréal, Montréal, QC, Canada
e-mail: mickael.trochet@umontreal.ca

N. Mousseau (✉)

Departement de physique and Regroupement québécois sur les matériaux de pointe, Université de Montréal, Montréal, QC, Canada

Department of Physics, University de Montréal, Montréal, Canada
e-mail: normand.mousseau@umontreal.ca; nm@normandmousseau.com

L. K. Béland (✉)

Department of Mechanical and Materials Engineering, Queen's University, Kingston, ON, Canada
e-mail: laurent.beland@queensu.ca

G. Henkelman (✉)

Department of Chemistry and the Institute for Computational and Engineering Sciences, University of Texas at Austin, Austin, TX, USA

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX, USA
e-mail: henkelman@utexas.edu

**Abstract**

Exact modeling of the dynamics of chemical and material systems over experimentally relevant time scales still eludes us even with modern computational resources. Fortunately, many systems can be described as rare event systems where atoms vibrate around equilibrium positions for a long time before a transition is made to a new atomic state. For those systems, the kinetic Monte Carlo (KMC) algorithm provides a powerful solution. In traditional KMC, mechanism and rates are computed beforehand, limiting moves to discretized positions and largely ignoring strain. Many systems of interest, however, are not well-represented by such lattice-based models. Moreover, materials often evolve with complex and concerted mechanisms that cannot be anticipated before the start of a simulation. In this chapter, we describe a class of algorithms, called off-lattice or adaptive KMC, which relaxes both limitations of traditional KMC, with atomic configurations represented in the full configuration space and reaction events are calculated on-the-fly, with the possible use of catalogs to speed up calculations. We discuss a number of implementations of off-lattice KMC developed by different research groups, emphasizing the similarities between the approaches that open modeling to new classes of problems.

# 1    Introduction

Modeling the dynamics of chemical and material systems is a fundamental challenge for computational scientists. While the equations of motion of atomic scale systems have been known since the days of Newton, their integration over experimentally relevant time scales still eludes us even given modern computational resources. With empirical potentials, we can achieve nanoseconds of simulation time per day, and with ab initio methods, we are limited to picoseconds. But in most chemical and material applications, we are interested in the human time scales of seconds to minutes, which are relevant for applications including catalysts and batteries. Bridging the time scale gap between what we can model with molecular dynamics (MD) and practical applications is key to making molecular simulations relevant.

Fortunately, many systems in chemistry and material science are what can be described as rare event systems. In this case, atoms in the material vibrate around equilibrium positions for a long time before a transition is made to a new atomic configuration, or state. This might be the hopping of a Li atom between sites in a battery material or a bond-breaking event in a catalytic reaction. With microscopic kinetics taking place on time scale well separated from phonons, rare event systems open the door to the application of various numerical solutions based on the transition state theory that are not available generally: instead of modeling the vibrational dynamics, which occurs on a femtosecond time scale, transition state theory can be used to average over the thermal motion and calculate the rate of the slowest rare event of interest. If all such important rates can be found, the state-to-state evolution of the system can be calculated on the time scale of the rare events.

A powerful approach for modeling the evolution of a system when the rates are known is the kinetic Monte Carlo (KMC) method. In traditional KMC, the mechanism and the rate of every possible event are required before a calculation is started. This requirement essentially limits atomic KMC simulations to be defined on a lattice where interactions between atoms and atomic motion can be defined discretely. In KMC, a table of all possible events is made, and a single event is chosen with a probability proportional to its rate. The amount of time that evolves between KMC steps is, on average, given by the inverse of the sum of the rates. Thus, each KMC step can be accomplished with just a couple of random numbers and the bookkeeping required to keep track of the possible events; this computational efficiency allows for KMC simulations of millions or billions of events and time scale orders of magnitude longer than any single elementary event.

Obvious limitations of traditional KMC are (i) the need to know all possible events a priori and (ii) the representation of atomic configurations on a lattice. Many atomic systems of interest, however, especially in the presence of defects or disorder, are not well-represented by a lattice-based model. Additionally, materials often evolve with complex and concerted mechanisms that cannot be anticipated before the start of a simulation; it is the evolution of the system into unanticipated configuration via unexpected events that make simulations most interesting.

The purpose of this chapter is to describe a class of algorithms, called off-lattice or adaptive KMC, which relaxes both limitations of traditional KMC. Specifically, atomic configurations are represented in the full configuration space, and reaction events are calculated on-the-fly so that the KMC event table is not fixed but rather adapts as the simulation progresses. Thus, off-lattice KMC can be applied to a much wider range of interesting systems, and the calculated evolution of the system can reveal unexpected dynamics that were not anticipated by the modeler.

There are now a number of implementations of off-lattice KMC developed by different research groups. Later in the chapter, we will discuss some of the differences in philosophy and specific algorithms, but a primary objective here is to highlight what is common between off-lattice KMC methods and emphasize that what may appear to be different methods with different names are in fact often minor variations on a common theme.

## 2    Off-Lattice KMC

Off-lattice KMC was inspired by a number of closely related earlier methods. In an approach by Sørensen et al., high-temperature MD simulations were used to determine possible reaction mechanisms between a probe tip and a metal surface (Sørensen et al. 1996). These pathways were refined using the nudged elastic band (NEB) method (Jónsson et al. 1998; Henkelman and Jónsson 2000; Henkelman et al. 2000) to determine the energy barriers and rates at the lower temperature of interest. While the authors did not fully sample a rate table for KMC, their work anticipates the philosophy of off-lattice KMC and, in fact, was later developed into an accelerated molecular dynamics method called temperature-accelerated dynamics (Sørensen and Voter 2000). A second method with close connections to off-lattice KMC is the activation-relaxation technique (ART) (Mousseau and Barkema 1998b) in which local arrangements of atoms were driven away from their minimum positions on the potential energy landscape, over ridges, and into neighboring basins of attraction. Again, no KMC event tables were calculated, but the idea of sampling reaction pathways from local displacements of atoms became a key idea for off-lattice KMC methods. ART was later extended into an off-lattice KMC method called kinetic ART, the details of which will be discussed here (El-Mellouhi et al. 2008; Béland et al. 2011).

The earliest implementation of off-lattice KMC that we know of was in 2001 by Henkelman and Jónsson (2001). In this method, local atomic displacements were made to initiate minimum-mode following saddle point searches; these searches iteratively find the lowest mode of the Hessian and follow this mode to a saddle point.

Over the 15 years since the first introduction of off-lattice KMC, there have been a number of significant improvements to the algorithm. Much of the development has been devoted to identifying and implementing efficient methods for storing and reusing information about reactive events that have been calculated. Another key issue, which will be discussed later, is the so-called low-barrier or flickering problem in which fast events limit the overall time scale that can be achieved by off-lattice KMC. Today, there are a number of approaches associated with different aspects of off-lattice KMC methods. Yet, they can generally be described in the single framework outlined in Fig. 1. In brief, all off-lattice KMC methods start from an initial state and explore configuration space around the state to find reactive events which lead to adjacent states. While not absolutely essential, existing methods tend to focus on finding saddle points on the boundary of the initial state and use the harmonic approximation to TST to calculate the rate of each possible reaction mechanism. Current saddle points can be found using a number of ways: minimum mode following saddle point searches, high-temperature MD, the recycling of saddle point information from previous steps, or informed by a database of topologies or structures. As new reaction mechanisms are found, they are added to the rate table and passed to any database being used to store event information. Also, as the configuration space around the initial state is searched for possible reaction events, and the KMC event table is constructed, an estimation of

**Fig. 1** Flow chart showing the general structure of off-lattice KMC algorithms

Initialize

Search for saddle points leading to neighboring states

- minimum-mode following methods
- high temperature molecular dynamics
- recycling saddles from previous steps
- database of local topologies
- database of local structures

New reaction mechanism found?

no

yes

Calculate rate and update kinetic Monte Carlo (KMC) event table

Update any database of kinetic events

Update confidence in the fraction of the KMC event table found

Is the KMC event table sufficiently accurate?

no

yes

Choose event to new state and update the simulation time according to the KMC algorithm

Is the desired simulation time reached?

no

yes

Stop

the completeness of the event table is updated. When there is sufficient confidence in the event table for escaping the current state, a KMC move is made to a new state, and the process is repeated.

The various off-lattice KMC implementations differ in the specifics in how saddle points are found and stored for reuse and how low barriers are managed. A number of the different philosophies and methods will be discussed next.

# 3 Search for Saddle Points Leading to Neighboring States

Even for relatively simple systems such as diffusing interstitials in metals, the complexity of the energy landscape is such that it is not possible to identify by hand all mechanisms (Henkelman and Jónsson 1999; Marinica et al. 2011). Open-ended event search methods are therefore necessary to identify the diffusion mechanisms and their associated barriers. There are two main classes of open-ended methods: (1) minimum-mode following methods and (2) molecular-dynamics based methods. Furthermore, local structures and topologies can be classified in order to identify where saddle points searches should be performed and also in order to classify and reuse these events in subsequent steps.

## 3.1 Minimum-Mode Following Methods

The minimum-mode following approach stabilizes first-order saddle points by finding the direction of lowest curvature on the potential energy surface, inverting the force in that direction, and relaxing the system, guided by these modified forces. A number of such methods have been proposed. The method used by Henkelman and Jónsson was named the dimer method because two images, separated by a finite displacement, were used to approximate the local curvature (Henkelman and Jónsson 1999). The dimer method was later shown to be equivalent to a method developed at the same time in the Wales' group, called the hybrid-eigenvector following method (Munro and Wales 1999). In fact, both methods use a Raleigh-Ritz quotient for iteratively determining the lowest curvature mode and a force inversion along this mode to stabilize first-order saddle points. Around the same time, ART was proposed, using a force projection approach to move the conformation toward saddle point (Barkema and Mousseau 1996). For better stability and convergence, the force projection was replaced by a Lanczos-based algorithm a few years later, forming ART nouveau (ARTn) (Malek and Mousseau 2000). It was subsequently further optimized, as described in details in Machado-Charry et al. (2011). Through benchmarking (Chill et al. 2014a) and a mathematical analysis (Zeng et al. 2014), it is now clear that these minimum-mode following methods are essentially equivalent in terms of computational efficiency for finding saddle points and that they differ mostly in their specific implementation.

A typical minimum-mode following search occurs in three steps:

1. From a local minimum, an atom and possibly its neighborhood are displaced in a direction that can be initiate randomly or systematically which aim to push the system out of the local harmonic state. To avoid collisions between atoms, the system can be partially relaxed in the hyperplane perpendicular to the displacement. This procedure can also be repeated until the lowest eigenvalue becomes negative or falls below a given threshold (typically between $-1$ and $-10 \, \text{eV/Å}^2$ for bulk semiconductors and metals).

2. The system is iteratively maximized along the negative (lowest) eigenvector and minimized in the hyperplane orthogonal to this direction. In the dimer and hybrid-eigenvector following methods, this is accomplished by following an effective force with the component along the negative mode inverted. Any optimizer, such as L-BFGS (Nocedal 1980), can then be used to converge to a first-order saddle point. In the ARTn method, the system is moved along the direction of the negative eigenvalue away from the initial minimum, and the energy is minimized in the orthogonal hyperplane at each iteration. In ARTn, if at any point the lowest eigenvalue becomes positive, iteration is stopped, and a new event search is launched, going back to (1). If the lowest eigenvalue becomes positive in the dimer method, the system is pushed up along the lowest positive mode until a negative mode is recovered. For either of these methods, a saddle point is considered located when the total force on the system falls below a given threshold (typically 0.01 eV/Å) with a negative lowest eigenvalue.
3. From the saddle point, the system is displaced along the negative mode and relaxed to find the connecting final minimum, completing the event.

Although the initial deformation is often limited to an atom and its surroundings, all the atoms in the box are allowed to respond to this change and to move, avoiding constraints on the nature and the size of the transition state. As the system leaves the harmonic basin and converges onto a first-order saddle point, however, many atoms initially displaced fall back close to their original position as events tend to be local in nature. In ARTn typically 50% of the time the negative eigenvalue is lost as the system is pushed along the eigenvector associated with the lowest eigenvalue, and it falls back into the initial minimum, as the structure of the energy landscape can include shoulders and bumps that do not lead to transition state. Any attempt to eliminate these lost events requires large initial deformations that typically bring the system to saddle points that are not directly connected to the initial minimum, breaking the continuity of the trajectory. With the dimer method, when a negative mode is lost, the system follows the lowest mode up the potential until a negative mode is recovered. While this approach allows minimum-mode following searches to push through positive curvatures, there is a greater chance of finding saddles that are disconnected from the minimum. The recently developed $\kappa$-dimer method, (Xiao et al. 2014) which uses the isocontour curvature to detect boundaries of the initial state, largely eliminates the problem of disconnected saddles.

The discovery of disconnected saddles is not always a problem. Wales and collaborators, for example, use large deformations to facilitate the construction of a transition matrix and can be then used to extract global kinetic information (Wales 2002). For other applications, however, it is essential to produce a continuous trajectory, and all transition states are tested to ensure that they are directly connected to the initial minimum.

As long as the lowest eigenvalue remains negative, moving to the saddle point is straightforward. While a convergence force of 0.01 eV/Å is generally chosen, ensuring convergence of the energy barrier to less than 0.01 eV within a few hundred force evaluations, the convergence criterion can be tightened.

## 3.2    High-Temperature Molecular Dynamics

An alternative to minimum-mode following methods for finding saddle points is to perform high-temperature MD initiated within the current state and use periodic quenching to see if a transition has been made. When a transition to a new state is detected, a double-ended saddle search method is used to find the minimum energy path between the two states. An efficient strategy for this is to use the climbing-image nudged elastic band method (Henkelman and Jónsson 2000; Henkelman et al. 2000) and then optionally, to save computational time, switch to a minimum-mode following search from within the neighborhood of the saddle.

The MD approach for finding saddle points from an initial state is very similar to the temperature-accelerated dynamics method (Sørensen and Voter 2000). The difference is that temperature-accelerated dynamics uses high-temperature MD to find for the first escape that would take place at the low temperature of interest, using harmonic TST to extrapolate to the low-temperature escape time. For off-lattice KMC, high-temperature MD is used to find all of the low-energy escape mechanisms and rates. What is common to both methods is that the use of MD to search for saddles can provide a rigorous confidence measure in the accuracy of the simulation. In the case of off-lattice KMC, this confidence measure is the fraction of the rate table which has been found using the MD saddle searches (Chill and Henkelman 2014).

Each MD saddle search typically takes more computational work to find a saddle than a minimum-mode following search. For off-lattice KMC, however, the cost of finding any one saddle is less important than the cost of evaluating the full rate table. For that, MD and minimum-mode following searches are competitive. If there is sufficient knowledge about the system to target minimum-mode following searches, they can find the rate table more efficiently. However, if there are many processes available to the system, MD searches can selectively find those which are most important because events with a higher rate are found more rapidly than slow events with MD searches (Chill and Henkelman 2014). Independent of efficiency, however, the main benefit of the MD searches is the confidence measure that it provides and the simplicity of having just one parameter to set, the high temperature, although anharmonicity may induce a false sense of completeness as discussed below.

## 4    Classifying Local Off-Lattice Environments

It is a good practice to classify local structures in order to identify those on which the saddle point searches should be centered, as well as to catalog events to be reused for future use. Several strategies for classifying structures and storing saddle point information are described here.

## 4.1    Lattice-Based Classification

To be useful, structural classification must be able to reduce a wide range of global conformations onto a finite set of locally defined variables. When atomic motion is limited to a discrete set of positions, there is no ambiguity when comparing two states: they're the same or they're not. For off-lattice calculations, atomic positions can occupy a continuous range of values. In effect, it is highly unlikely for two different global conformations to present identical local environments. To classify and compare structures in an off-lattice system, it is possible to define a discretization procedure that will map the continuous array of solutions into a discrete set of states while ensuring that these states share sufficient similarity when it comes, for example, to their list of diffusion mechanisms.

Pattern recognition approaches, such as the one adopted by Trushin et al. (2005) with the self-learning kinetic Monte Carlo method (SLKMC), are a step away from standard predetermined catalogs, allowing to treat a wider range of conformations and, therefore, better include local strain effects. Nevertheless, this algorithm still requires an underlying lattice to ensure the discretization of the configurations and events, and, while barriers are constructed on the fly, the event catalog consists only of single-atom nearest-neighbor hops characterized using the drag method.

With the self-evolving atomistic kinetic Monte Carlo (SEAK-MC), Xu et al. (2011) propose an off-lattice approach limited to near crystalline configurations with an on-the-fly event searching step. A particular feature of this approach is the construction of *active volumes* associated with defects or noncrystalline environments. These defects are identified using a geometric criterion. Focusing on these defective regions reduces the computational effort for constructing an event catalog. Dimer (or ARTn Béland et al. 2015a) searches are used to search for mechanisms, starting from each defect in the active volume. After a KMC step, only the defects in the affected active volume are sampled for diffusion; events associated with other regions are simply carried over from the previous step. Although SEAK-MC can handle more complex environments than SLKMC, because of its reliance on defects to identify active volume, the method is not applicable to fully disordered configurations. Since it does not reconstruct all barriers after each step, moreover, SEAK-MC does not fully take into account long-range elastic effects.

## 4.2    Topological Classification

Introducing a topological classification, the kinetic activation-relaxation technique (k-ART), first published in 2008 (El-Mellouhi et al. 2008), lifts these limitations and is the first fully off-lattice KMC approach with on-the-fly cataloging.
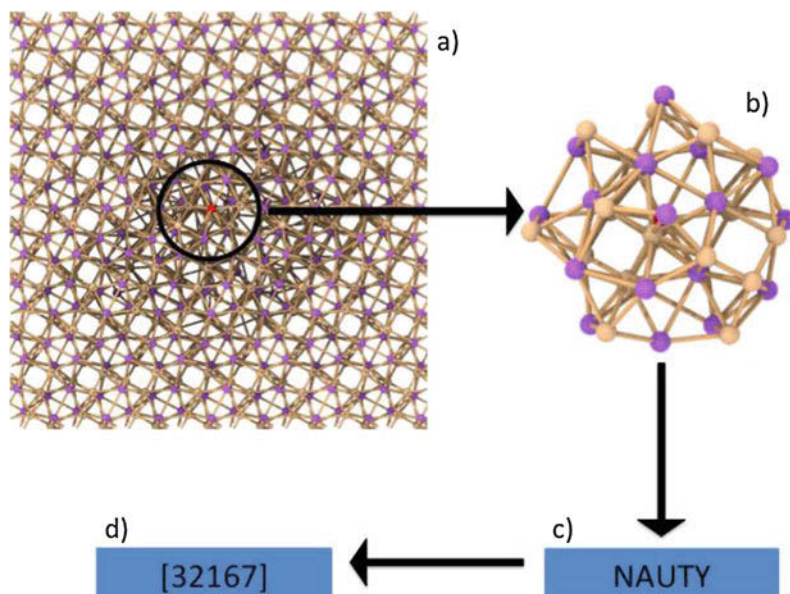
**Fig. 2** The topological classification of each atom, in k-ART, is determined by its local environment. (**a**) An atom and its neighborhood are first identified; (**b**) vertices are drawn between atoms within a cutoff distance generally selected between first and second neighbor; (**c**) the resulting graph is then sent to NAUTY, which returns an identifier characterizing the automorphic class its belongs too and a set of transformation onto a reference graph

K-ART attributes a topological key to each atom in a cell based on its local environment (see Fig. 2): all neighboring atoms within a sphere centered on the reference atom are considered as vertices of a graph, with edges drawn between atoms within a specific cutoff, typically, but not always, set between first and second nearest distances. As a function of the system's complexity, the sphere radius is generally set between 6 and 8 Å, including around 50 to 80 atoms. The graph, as generated, is then sent to NAUTY, a rich topological analysis code developed over many decades by McKay (McKay et al. 1981; McKay and Piperno 2014) used as a library by k-ART. NAUTY returns a key characteristic of the graph's automorphic class as well as a mapping list into a reference graph. This list is used by k-ART to map specific local environments onto reference configurations found in the event catalog.

For each key, a series of ART nouveau event searches is launched centered on the atom associated with this key. Saddle points with different topological keys are considered unique and stored. The topological key associated with a specific atom defines the list of possible events it can undergo. Using the mapping list, it is possible to assign a correspondence between the specific environment and the atoms in the catalog associated with the generic event. Using this correspondence and the

atomic positions of the generic configuration at the saddle point, one can reconstruct an approximate saddle point for the specific configuration. In the presence of strain, this saddle point will be close to the reference event, but not exactly the same, and a few steps of ART nouveau relaxation are needed to converge onto the specific saddle point, providing a precise information on the geometry and energetics of the transition state.

K-ART's basic assumption poses that local minima in the same automorphic class share a unique set of events, characterized by the topological classifications, with the geometrical details at the saddle point depending on the elastic deformation. This allows k-ART to store events based on the local topology and reconstruct their specific geometry at these specific points on the energy landscape. This assumption is valid most of the time for three reasons: (i) the correspondence is set to work only at specific points on the energy landscape – minima and first-order saddle points; (ii) the correspondence holds therefore only for a given forcefield or, equivalently, a unique energy landscape; and (iii) the local graph is reconstructed after it is embedded within the larger two- or three-dimensional space and attached to the rest of the network. This one-to-one correspondence between topology and geometry can fail. However, this failure can be identified readily as activated configurations reconstructed from a reference geometry will not show a first-order saddle point in their vicinity. In general this error indicates that more than one geometries is associated with the topology. To lift this degeneracy, k-ART reduced the edge cutoff criterion until these geometries are assigned to different classes.

PESTO, by Louis Vernon (Vernon et al. 2011; Vernon 2010, 2012), is a variation on adaptive KMC, which has integrated a number of k-ART's features. While topological classification in k-ART is based on the local environment of *atoms*, PESTO is based on the local environment of *defects*. This is a three-step process. First, defects are identified using a variety of schemes including comparison to a reference lattice. These defects can be point or extended defects. Second, NAUTY is used to identify the defect's topology. While k-ART looks at cluster of atoms centered on individual atoms, PESTO looks at cluster of atoms centered on defects. Third, the positions of the atoms in this cluster are compared to those of previously stored clusters for this topology. If they match those of the stored cluster, the configuration is considered to be known. If they do not, the configuration is considered to be new and it is stored.

In 2016, Alexander and Schuh developed a version of k-ART that used a systematic and orderly search and classified transition based on atomic motion vectors, rather than a topological classification. Displacement vectors of each atom between the saddle point and the initial state and between the final and the initial states are stored for future comparisons. Comparison of two transitions is performed by comparing the $x$, $y$, and $z$ displacement of each atoms of both transitions. The goal pursued by Alexander and Schuh is to evaluate the convergence of the residence time by assessing the completeness of search for any configuration in the potential energy landscape.

## 4.3    Geometric Classification

An alternative to characterizing local environments by the bonding topology is
to directly use the atomic geometry of the atoms which participate in a reactive
event. In the kinetic database (KDB) approach (Terrell et al. 2012), all atoms
that move by more than a specified distance are considered part of the reaction
mechanisms. The position of those atoms in the initial, saddle, and final states, as
well as their direct neighbors, is stored in the KDB. Figure 3 shows a calculation
of CO oxidation on Au(111) in the presence of hydroxyl, using forces from density
functional theory (Ojifinni et al. 2008). When a hydroxyl formation mechanism is
discovered, the local geometry of the reacting atoms are stored in the KDB. Later
in the simulation, or in an entirely different simulation, the KDB tries to match the
reaction mechanisms stored to the current geometry. Specifically, matching is done
using a geometric fingerprint of each atom in the database structure (initial or final
state geometries). Using a depth-first tree search, neighboring atoms in both the
database and current configuration are matched. If a complete match is made, the
optimal alignment is calculated, and a score is assigned to each candidate, based
upon the difference in atomic positions between those in the database and those in
the current configuration. In Fig. 3, the final hydroxyl structure in the database was
matched to a hydroxyl on the surface in a subsequent step in the off-lattice KMC
simulation. The KDB is then used to predict possible saddle points by moving the
local atoms to their saddle point configuration from the database. These suggested
saddle geometries are used as initial configurations for minimum-mode following
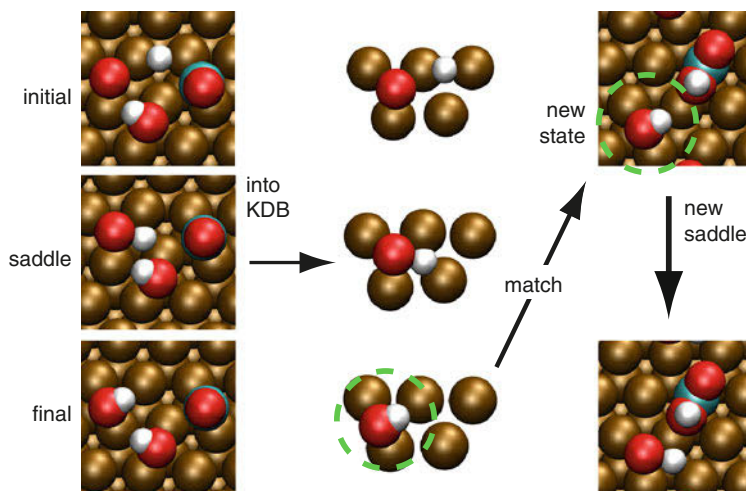


**Fig. 3** Kinetic database example of OH and CO reacting on Au(111). When a new reaction
mechanism is found, the configurations of the moving atoms and their first neighbors are stored in
the database. In any subsequent state, the database is queried to see if any known initial or final
states match

searches so that the true saddle point and rate of the event in the new environment are calculated. Importantly, the KDB is used to provide approximate saddle point configurations for the current geometry based upon what has been seen before. Only if an exact saddle can be found is the mechanism entered into the rate table. In this way, there is no additional approximation for off-lattice simulations using the KDB. As the database stores a greater number of kinetic events, saddle suggestions become more accurate and converge more rapidly to true saddles, when they exist, so that the cost of the off-lattice KMC simulation approaches that of KMC.

There are some key differences between the topological classification of reactive events used in k-ART and the geometric classification in the KDB. K-ART has the advantage that matching structures from the database to the current simulation is extremely efficient. Also, if there is sufficient trust in the catalog of events in the database, then new saddle searches need only to be done when an unknown topology is reached. In the KDB, on the other hand, the structure matching is not discretized; geometries which are close, but not exact, are used to suggest saddle point structures. When these structures are converged, the exact geometry of the saddle and the activation energy are calculated. If no saddle point is found, the loss is only the computational time required for the saddle search. Both methods can be used to store and recover information that has been learned to accelerate off-lattice simulations.

## 4.4     Challenges in the Event Reconstruction

Reconstruction of saddle points is crucial in any scheme that creates a generic catalog built either on geometrical or topological keys as symmetry operations and elastic deformations in an off-lattice system must be taken account.

To describe this challenge, we focus here on the topological approach used in k-ART, which creates a catalog based on reference geometries at the initial minimum and saddle point and uses the knowledge of current geometry at the minimum and the correspondence between the reference and the current topologies at the minimum to reconstruct a new saddle point.

1. Atoms of the current minimum configuration are matched with the reference minimum using the correspondence between the two graphs drawn by NAUTY.
2. A reference frame is then constructed, by comparing the relative positions of the various atoms in the two configurations.
3. Atoms in the current minimum are then displaced using this reference frame, according to the difference between the reference saddle point and energy minimum.
4. The energy and forces are evaluated on the reconstructed activated state. If the activation energy difference between the reconstructed and the reference state and the absolute forces in the reconstructed state are below given thresholds, the reconstruction is considered successful ARTn is applied to converge to the nearby saddle. If not, then the possibility that many correspondences between the

reference and current graphs exist is considered and a number of permutations in the correspondence labeling are attempted, based on various reconstructed reference frames.

5. When no attempted reconstruction is accepted, k-ART concludes that the correspondence between topology and geometry for this environment is not unique. The cutoff for defining edges between atoms is then modified, and the topology is split. If a new topology is found in the way, it is then populated by events, if a known topology is generated, the algorithm starts over for generating a specific event (Fig. 4).

The success in reconstruction depends closely on the graph. For compact systems, such as bulk metals or semiconductors, a graph containing between 50 and 70 atoms, with edges drawn between nearest neighbor atoms, is sufficient. For an anisotropic system such as graphene, however, it is necessary to define long-enough cutoff to ensure that planes are linked. Reconstruction also requires that a specific saddle point related to the generic one exist. For high-enough barriers, typically above 0.1 eV, a deformation strong enough to make such a barrier disappear would lead to a change in topology. For low barriers, particularly surrounding unstable points, this is not always the case. To avoid this problem, one can systematically recreate the event catalog associated with a given topology when such a low-energy barrier disappears, either automatically cross or simply ignore them.
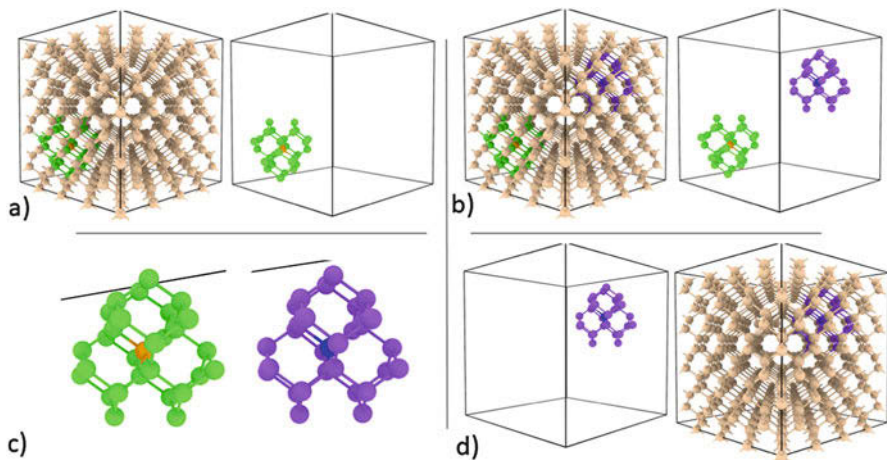


**Fig. 4** Reconstruction of a saddle point from the topological classification. (**a**) ART nouveau generates an event from the initial topology of a central atom (red sphere). (**b**) Once ART nouveau search is done, k-ART reconstructs each event associated with the same initial topology: topological correspondence of two central atoms (blue and red spheres). (**c**) Applying geometrical symmetry is needed to the initial structure of the second central atom (blue sphere) as well as the saddle point and final structures. (**d**) Finally, guess structures (i.e., saddle point and final applied of the second central atom) are then refined using ART nouveau to include elastic deformations

# 5  Confidence in the Completeness of the Saddle Point Catalog

An important parameter for off-lattice KMC is determining how many saddle point searches must be done. Too few searches will lead to incorrect kinetics, but searches can be computationally wasteful if all important transitions are already known. Some methods – such as k-ART and SEAK-MC – perform a finite number of searches for each defect that is encountered. In the case of k-ART, the event catalog is reused for all know topologies; new searches are performed to complete the catalog or to complete it, if a new topology is added. New searches are also performed to improve the event catalog, for very common topologies (searchers are added every log-10 times a topology is found). In SEAK-MC, the event catalog for a given defect is destroyed after the execution of one of those events and is rebuilt from scratch every time a defect is encountered.

When using minimum-mode following searches to find saddles, the completeness of the KMC catalog can be estimated from the statistics of how often new saddles are found. For example, if an increasing number of searches are done without finding any new events, there should be a growing confidence that the rate table is complete. Importantly, however, there can be a wide range in the frequency at which different saddles are found, and this bias can lead to unquantifiable uncertainties in the completeness of the catalog (Henkelman and Jónsson 1999).

Another way to estimate the completeness of the event table is to use MD saddle searches, as discussed previously. The key difference between MD searches and minimum-mode following searches is that the probability of finding any event using MD is proportional to its rate. Since we are interested in rare events, basin-constrained MD at an artificially high temperature is used to determine possible reaction mechanisms, and harmonic TST is used to calculate the rate of the events at the low temperature of interest. To the extent that harmonic TST holds, one can calculate a well-defined estimator for the fraction of the rate table that is missing, based upon the amount of MD time that is used to determine the rate table and the events that have been found (Chill and Henkelman 2014).

It should be noted that while high-temperature MD searches can provide an estimate of the completeness of the catalog, the high-temperature event catalog may miss events for which the rate decreases as temperature increases. Such non-Arrhenius behavior would be due to strong entropic effects. Examples of such behavior include thermal stabilization of nano-voids (Perez et al. 2013) and dislocations (Kim and Tadmor 2014).

# 6  The Low-Barrier Problem and Coarse Graining

While off-lattice KMC is able to coarse grain over the fast atomic vibration time scale and model the slower time scale of the state-to-state dynamics, it is ubiquitous to have another separation of time scales between the fastest state-to-state events and

slower time scales of interest. This so-called low-barrier problem describes how off-lattice KMC simulations can spend all of the computational time on fast events so that they are not able to reach long time scales.

A key idea that has been used to bridge the gap between fast and slow events is the framework of the Monte Carlo with adsorbing Markov chains (MCAMC), which was described by Novotny in 1995 (Novotny 1995). In this pioneering work, Novotny showed how the kinetics of a system that was partitioned into a set of transient states and adsorbing states could be solved to give a specific time for the transition from a transient state to a specific adsorbing state. Remarkably, this was shown to be possible without adding any approximations beyond the rates that are used in KMC regardless of the partitioning between transient and adsorbing states. This means that in an off-lattice KMC simulation, any set of states can be identified as transient, and the time to transition to a neighboring adsorbing state can be calculated using the MCAMC approach without additional approximations.

The exact calculation of transition times from transient states to adsorbing states requires an iterative set of matrix calculations each with dimension of the size of the total number of states involved. Novotny also showed that it was possible to simplify this calculation and instead calculate the average escape time from the transient states via the first moment of the escape time distribution (Novotny 2001). This approximation was later referred to as the mean-rate method (MRM) by Puchala et al. (2010) since the escape time from the set of transient states is characterized by a single rate. The mean-rate approximation works well when there is a separation of time scales between the transient and the adsorbing states. In this regime, the set of transient states has been referred to as a superbasin. The mean-rate method reduces the computational cost of escaping a superbasin to a single matrix inversion.

The MCAMC approach allows off-lattice KMC simulations to switch, at any point, between a KMC description of the state-to-state dynamics to a transition between any specified state of transient states to the neighboring adsorbing states. When MCAMC is done exactly, any set of transient states can be chosen, and so different definitions of the set of transient states have been proposed. On one extreme, all visited states can be considered transient so that every transition is made to a new state. This strategy was described in the Markov web, proposed by Boulougouris and co-workers (Boulougouris and Frenkel 2005; Boulougouris and Theodorou 2007). While advantageous in terms of being able to visit new states as rapidly as possible, this comes at the cost of losing the state-to-state detail of the trajectory to states that have been visited. Additionally, for simulations which are out of equilibrium and explore new parts of configuration space, including all states in the transient space add unnecessary cost to the matrix operations. Another possibility is to count the number of times that each state has been visited and combine states into a superbasin when the visit-count exceeds a specified threshold (Chill et al. 2014b). It is also possible to consider the energy of the states and the saddles between them (Pedersen et al. 2012) or combine states into a superbasin when they are connected by rates above a specified value, as in the basin-autoconstructing MRM (bac-MRM) (Brommer et al. 2014). However the superbasins are defined, they can grow as more transient states are explored.

# 7        How to Reduce Computational Costs

The main computational costs of a typical on-the-fly KMC run are associated with the evaluation of forces and energies and the construction of neighbor lists. Neighbor-list costs can be decreased using standard cell and Verlet neighbor list methods. Computational cost reduction, for force evaluations, requires taking advantage of the fact that the reference states, in KMC, are effectively at zero temperature, allowing algorithms to exploit the local nature of activated events. Given the large literature on the neighbor-list cost-reduction techniques, this section will focus on the forces.

A number of studies (Mousseau and Barkema 1998b; Pedersen and Luiser 2014; Xu et al. 2015; Guterrez et al. 2016) found that initial deformations applied to search for saddle points should be local in order to successfully lead faster to a diverse set of activated events. No more than a few hundred atoms should initially be involved in the event search, and usually less. As the positions of atoms are optimized while reaching the saddle point, more atoms will be deterministically displaced (typically, a few thousand); an even larger number will be displaced as the system is relaxed into a new minimum. Notably, extended cascade-type concerted events can take place during this phase (Béland and Mousseau 2013). In other words, while relaxation should be performed globally, activation can generally be performed locally to maximize the probability of successfully finding saddle points.

Search algorithms can exploit the locality of activated events (Mousseau and Barkema 1998b). The simplest way is to impose activation volumes. For instance, Xu et al. (2011, 2015) showed that a spherical active volume of 4 lattice parameters is sufficient to capture the activation barrier for vacancy diffusion in BCC iron within 0.001 eV, while an active volume of 6 lattice parameters was sufficient for the case of the dumbbell interstitial, and an active volume of 7.5 Å could capture the kinetics of 37-interstitial clusters (Xu et al. 2013). In other words, saddle searches can yield accurate results by calculating forces and energies over a subset of a few hundred atoms. It is also possible to find saddle points using small active volumes and then refine them using increasingly larger ones.

Imposing strict active volumes does have drawbacks. For instance, such an approach imposes static long-range elastic interactions, preventing a dynamical reaction of the whole system. To avoid this bias, it is possible to dynamically update a list of active atoms during the event search (Béland et al. 2011; Joly et al. 2012) defined by those atoms with a minimum force acting upon them and their neighbors. As the search progresses, an increasing number of atoms will have non-negligible forces. This procedure can effectively make computational time scale sublinearly with system size, with no strong assumptions to be made about the size of the active volume involved during event searches (Joly et al. 2012).

Such an approach, while elegant, is difficult to implement with standard force field libraries such as LAMMPS (Plimpton 1995). In this case, one turns to a partial use of active volume. After a global energy minimization of the whole structure, to ensure that all elastic effects are fully incorporated in the local minimum, a sphere is drawn around the central atom in the topology. An inner radius defines the set of
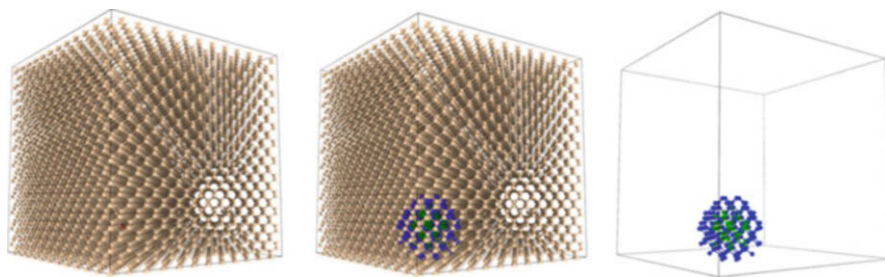
**Fig. 5** Graphical description of the local force calculations. Two spheres are centered on the same atom associated with a given topology. In the inner sphere (green atoms), all atoms can move following the activation. Atoms in the outer sphere (blue atoms) are fixed and contribute to ensuring that the correct forces apply to the inner atoms

atoms that are allowed to move during activation; a second, outer radius includes fixed atoms needed to ensure that the forces are accurate in the inner part. As a general rule, the inner sphere contains between 500 and 2000 atoms, while the outer shell is determined by the force field cutoff radius (Fig. 5) (Trochet et al. 2017). As with large enough active volumes, the error on the energy is small, and generic events are generated without further global relaxation at the saddle point nor the final minimum. Specific events, for their part, see by default the barrier fully relaxed at the saddle point after a first local relaxation, with the final configuration also relaxed globally. For systems with simple localized defects, it is possible to avoid a global relaxation at the saddle point for specific events, provided that the active volume is large enough, reducing the global relaxation only to accepted minimum states. The parameters to ensure a given error threshold must then be assessed specifically for each system. As a general rule, however, the use of local forces on systems containing 10,000 to 1 million atoms makes the event search almost order 1, considerably accelerating the algorithm without any significant loss in information or precision (Raine et al. 2017).

For the simulation of large systems, parallelization is crucial. An advantage of MC is that every event search can be performed independently, which means that many computer cores can simultaneously generate events to be added to the catalog. As mentioned in Sect. 7, one can accelerate the saddle searches by exploiting the locality of events. By increasing the number of "worker" computer cores proportionally to the number of saddle points to be generated, and by considering a local subset of atoms during saddle searches, the runtime between KMC steps will be approximatively independent of system size, if event management and KMC overhead costs are negligible.

# 8    Advantages of MC: Tricks and Shortcuts Available

Monte Carlo methods present a major advantage over molecular dynamics: it is much easier to use tricks to focus on the problems of interest, since the system's

evolution is event-based, instead of being continuous, offering a much better control on the rules that are used. In this section, we review a number of approximations, solutions, and tricks that make these off-lattice approaches more competitive that could be expected through a simplistic evaluation of their computational costs.

**The Low-Barrier Problem**  We have already presented the approach for handling processes with low barriers. Such methods are crucial for long-time simulations of complex materials, with many low-barrier events. By solving analytically the in-basin kinetics, it is possible to effectively perform billions of jumps without having to compute them directly, allowing the simulation to reach the relevant time scale instead of being slowed down to a halt by irrelevant mechanisms.

**Constant prefactor**  The use of constant prefactor is not essential, and some of us compute the prefactor for every event using the harmonic approximation. Nevertheless, for most compact system, the prefactor varies relatively little compared with the barrier, so it can be given a constant value, leaving the cost of evaluating a barrier to converging to a saddle point (Valiquette and Mousseau 2003). Clearly, however, as shown by Koziatek et al. (2013), the harmonic approximation fails in systems with significant density fluctuations, and it is generally necessary to, at least, demonstrate the validity of this approximation before using it. Prefactors can be affected by temperature, when barriers are low with respect to $k_{\mathrm{B}}T$, as discussed previously. Overcoming this limitation requires likely to turn to thermodynamical integration, which is computationally heavy and has not yet been automated.

**Restricting event searches to specific environments**  To decrease computational costs, it is possible to prevent off-lattice KMC calculations from searching for events in very stable environments. For example, when we are interested by a phenomena occurring on time scales of vacancy or interstitial diffusion, identifying events from perfectly crystalline environment, with barriers that are many eVs high, does not contribute to the dynamics. In some case, also, we can focus on specific regions, near a defect, for example, so that we may want to ignore the rest of the box, knowing that it will not contribute to the kinetics of interest. Off-lattice KMC calculations can therefore be instructed to ignore all crystalline topologies, specific atomic species, or regions when constructing its catalog.

**Biasing pathways**  While the previous shortcuts do not significantly affect the kinetics of the system on the time scale selected, it is also possible to give up the correct kinetics in exchange for exploring specific pathways. This can be done in many ways. First, a general bias in a given direction can be imposed, either through a selection bias from a complete event catalog or through the construction of an event catalog that only allows moves along a given pathway, specific mechanisms or a general direction. It is also possible to hand select, at every step, the event of interest among the list and evolve the system along a biased path.

   The list of tricks with Monte Carlo approaches is largely limited only by the researcher's imagination. In many ways, it is through these shortcuts that allow a

better focus on the important physics that KMC methods are most useful that, in addition to accelerating the simulation, they provide a much clearer picture of the fundamental mechanisms dominating specific processes.

# 9    Applications

Off-lattice KMC methods have been used to study a wide range of system, including metal on metal diffusion (Henkelman and Jónsson 2001), interstitials and vacancy clusters in c-Si (El-Mellouhi et al. 2008; Trochet et al. 2015), methanol decomposition on Cu, (Xu et al. 2009) Fe (Brommer et al. 2014; Restrepo et al. 2016) and Ni (Mahmoud et al. 2018), ion-implanted relaxation in Si (Béland et al. 2013; Béland and Mousseau 2013; Jay et al. 2017), ion- and neutron-irradiated metals (Béland et al. 2015b; Lu et al. 2016) and alloys (Béland et al. 2016; Lu et al. 2016; Osetsky et al. 2016), grain boundary diffusion in Cu (Pedersen et al. 2009), hydrogen diffusion in Al grain boundaries, (Pedersen and Jónsson 2009) a solid-solid phase transformation in Mo (Duncan et al. 2016) defects in amorphous Si (Joly et al. 2013), Li impurities in Si (Trochet and Mousseau 2017) and C impurities in Fe (Restrepo et al. 2016, 2017), and many more.

We review here a few applications that represent some of the strengths and limitations of these methods.

## 9.1    Loop Transformation in FeCr

Off-lattice KMC is a powerful tool to simulate the long-time kinetics of point defects and small defect clusters. However, it is not limited to these relatively simple problems and can be used to capture slow kinetics involving extended defects. A good example of such a problem is the transformation of $1/2\langle 111 \rangle$ prismatic dislocation loops in Fe into $\langle 100 \rangle$ prismatic dislocation loops. $1/2\langle 111 \rangle$ and $\langle 100 \rangle$ loops in bcc Fe and FeCr are common interstitial-type defects observed in neutron- and ion-irradiated samples. The $1/2\langle 111 \rangle$ loops are known to be fast one-dimensional diffusers, which are closely related to void swelling. $\langle 100 \rangle$ loops are largely immobile. The ratio of formation of $1/2\langle 111 \rangle$ to $\langle 100 \rangle$ loops is directly linked to void-swelling rates. However, collision cascades simulations in bulk Fe predict the formation of $1/2\langle 111 \rangle$ loops, but not of $\langle 100 \rangle$ loops. There must be a post-cascade mechanism that permits the transformation of the post-cascade defects into $\langle 100 \rangle$ loops. High-temperature MD simulations – more than $1000\,\mathrm{K}$ – of two $1/2\langle 111 \rangle$ loops intersecting and transforming into a single loop suggested that this reaction solely leads to one large $1/2\langle 111 \rangle$ loop (Terentyev et al. 2008). Using off-lattice KMC in Fe (Xu et al. 2013) and FeCr (Béland et al. 2015a), a novel mechanism for the transformation of two $1/2\langle 111 \rangle$ into a $\langle 100 \rangle$ was discovered, which is illustrated in Fig. 6. At $600\,\mathrm{K}$, the waiting time was $200\,\mathrm{ns}$ in off-lattice KMC – i.e., a $0.73\,\mathrm{eV}$ overall activation barrier. This mechanism was later confirmed by molecular dynamics, as reported in Xu et al. (2013). The simulations in FeCr
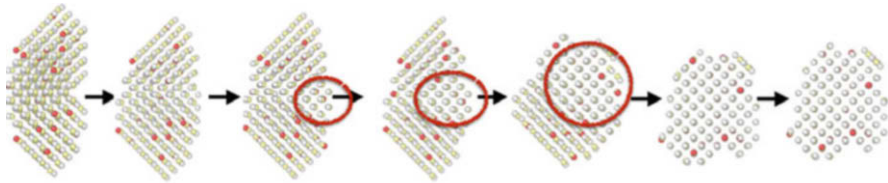
**Fig. 6** An illustration of the transformation of two $1/2\langle 111 \rangle$ interstitial clusters into a $\langle 100 \rangle$ loop in FeCr (10 at. % Cr). The red circles indicate the subregion that is changing orientation within the interstitial cluster. Gray spheres are Fe atoms not sitting on bcc lattice sites, orange spheres are Cr atoms not sitting on bcc lattice sites, and yellow atoms are empty bcc lattice sites. The simulation cell contains 16074 atoms. (The figure is adapted from Béland et al. 2015a)

(10 at. % Cr) also indicated that Cr decoration of the interstitial clusters favor the transformation to a large $1/2\langle 111 \rangle$, relative to the situation in pure Fe. This was a good example of a situation where off-lattice KMC was able to predict a mechanism that helped guide further MD simulations.

## 9.2    Phase Transformation in Mo

Figure 7 illustrates both the success and limitations of the off-lattice KMC method for the simulation of a transition between a complex A15 phase to the lower-energy BCC state in bulk Mo (Duncan et al. 2016). In this calculation, the interatomic interactions are described by an embedded atom method potential (Zhou et al. 2001). The formation and dissolution of complex (topologically close-packed) phases are important both for understanding the hardness and fracture of Ni-based superalloys (Sinha 1972). In Fig. 7a the phase transition from A15 to BCC is observed with AKMC to occur at 300 K over time scales of microseconds. The atoms are colored according to a common-neighbor analysis (Faken and Jónsson 1994) to better visualize the interface (gray) between the A15 (red) and BCC (blue) phases.

The simulation in Fig. 7a involves only a few hundreds of atoms, and yet we can already see a system-size-related problem caused by the disorder at the phase boundary. Specifically, as shown by the disconnectivity graph in Fig. 7b, there are many states connected to the initial state by low barriers which form a superbasin. All of these states must be enumerated in the AKMC simulation, as well as the rates between them. Only then a MCAMC move can be used to find a higher barrier process leading to interface motion, to the final state indicated. An analysis of the superbasin states shows that most transitions within the superbasin involve groups of atoms switching from A15- to BCC-coordinated. The fundamental problem is that when there are several such independent groups of flickering atoms, the total number of states in the superbasin grows combinatorially with the number of groups. This catastrophic situation is shown clearly in Fig. 7c where the supercell is increased and the total number of states in the superbasin is so large that even a single-layer transition between A15 and BCC cannot be observed with our off-lattice KMC machinery.
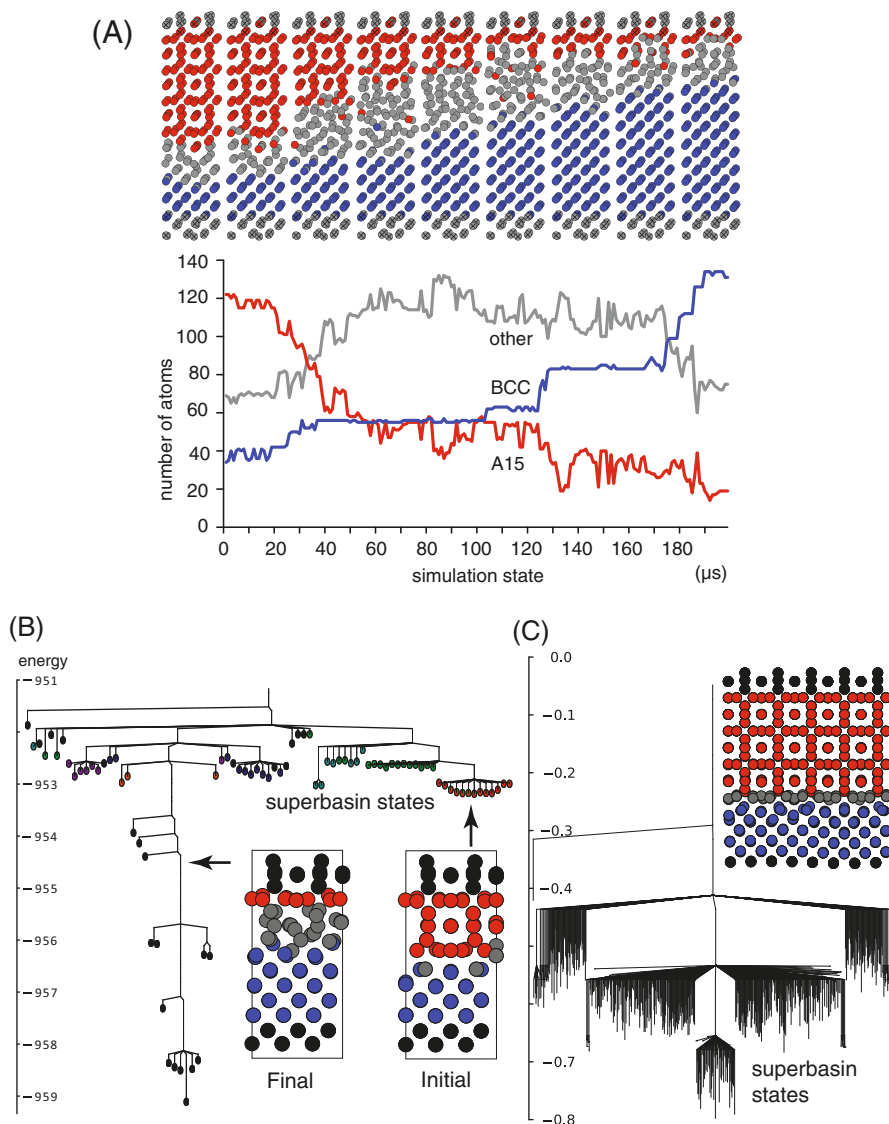
**Fig. 7** (**a**) AKMC simulation of a transition from the A15 complex phase in Mo to the lower-energy BCC phase, which occurs on the timescale of microseconds at room temperature. (**b**) A disconnectivity graph showing that disorder at the interface between the phases gives rise to many states connected by low barriers. (**c**) As the system is made larger, the number of states in the superbasins grows combinatorially, making the simulation intractable for simulating even a single-layer transition

## 10      Discussion

Off-lattice KMC approaches have allowed to study the long-time dynamics of systems that were long off-limits, including systems with many defects, alloys, interfaces, grain boundaries, and even fully disordered materials. These methods have also demonstrated that, even for systems that appeared simple, unexpected mechanisms could play a significant role and that even lattice-based methods would benefit from constructing their event catalog using unbiased search methods.

### 10.1     Limits

While open-end saddle point search methods such as ART, ART nouveau, and the dimer methods are powerful, we still lack a fundamental theoretical support for establishing the completeness of the generated event catalog. How many event searches should one launch by atom? When can one be certain that all relevant events are found? From experience, we know that, with ART nouveau, lower-energy barrier are found much more often than very high barriers. For a method such as k-ART, this is certainly positive. Moreover, comparison with MD and other methods, when possible, suggest that it recovers all previously identified mechanisms. Yet, it is not possible to assess, even statistically, what the error is with these methods contrary to what can often be obtained by the MD-based accelerated approaches of Voter and colleagues (see discussion above). Theoretical bounds or limits would certainly help greatly to ensure that these simulations do capture the essential physics.

While off-lattice with or without on-the-fly catalog building KMC approaches can reach time scales inaccessible to MD, these methods remain much heavier than lattice-based atomistic KMC and are typically limited to a tens of thousands of KMC steps, not counting, for course, analytically handled flickers. These methods can therefore only be applied to systems where physically relevant mechanisms are largely dominated by local mechanisms.

For example, vacancy-induced solute diffusion in metal, where one must wait for a vacancy to diffuse near a solute to see the atom jump, is much too expensive for these types of simulations. To overcome these limitations, one could solve directly in the equilibrium distribution of vacancies of the system for the time step between configurations with vacancies near a solute or couple off-lattice KMC with lattice-based KMC so that the vacancy diffusion away from solutes can be solved efficiently while the elastic effects and interactions between solutes and points defects are addressed exactly with off-lattice description. Clearly, other possibilities exist, and work will have to be performed in this field to assess the best way of coupling these scales.

Also, the total number of defects that can be effectively handled by off-lattice KMC is typically limited to a few hundred defects. Increasing the number of defects increases the computational cost in two ways. (1) Events must be assigned to each additional defect. While efficient recycling of events can minimize the

associated computational cost, such recycling is not always an option. For example, in concentrated alloys, the number of local configurations grows combinatorially with the number of elements, which limits the usefulness of recycling events. (2) As the number of defects increases, the total rates increases as well, which reduces the waiting time of each KMC step. More KMC steps – i.e., more computational resources – are necessary to reach the same time scales. While synchronous (Shim and Amar 2005; Martínez et al. 2008) and asynchronous (Shim and Amar 2006) KMC parallelization techniques offer a possible solution to this problem, this is still an unresolved challenge for off-lattice KMC.

Another fundamental limit of on-the-fly KMC is that it probes the potential energy surface. If the problem of interest involves significant entropic contributions to the free energy, exploring the potential energy surface might lead to incorrect predictions. On the other hand, many free energy landscape-based methods suffer from the "curse of dimensionality" (Althorpe et al. 2016); to be effective, these accelerated methods necessitate that a proper – and relatively small – set of reaction coordinates be inputed. In the future, we can imagine on-the-fly KMC being used to find promising reaction coordinates to be used as input for accelerated free energy methods.

## 10.2   Future Developments

Beyond these questions, a number of other developments should be undertaken to improve the efficiency of off-lattice KMC methods.

**Moving beyond master-worker parallelization**  As larger systems with more defects are simulated, the computational overhead of the master-worker parallelization scheme of current off-lattice KMC codes will become a bottleneck. The way forward may be to switch to a decentralized parallelization scheme. While it entails challenges in regard to cataloging, recycling, and load-balancing, such an approach could significantly increase the scalability of off-lattice KMC.

**Recycling basins to handle flickers**  The absorbing Markov chain algorithm described above – i.e. the superbasins – can be define states either as global configurations – cf. Sect. 6 – or local configurations – (Fichthorn and Lin 2013). The latter can provide a significant acceleration over the former as it can decouple the kinetics of non-interacting defects. This local treatment of superbasins is implemented in a number of the current off-lattice codes. A further acceleration would be to reuse superbasins built previously if the configurations are encountered again. This would increase the bookkeeping costs but can largely use the same geometrical and topological classification tools that are currently used to recycle events.

**Extended defects**  Off-lattice KMC can handle certain extended defects. For example, it was used to simulate the propagation of a grain boundary in Mo (Duncan

et al. 2016) and to simulate the transformation of prismatic dislocation loops in bcc Fe and FeCr (Xu et al. 2013; Béland et al. 2015a). However, the kinetics of dislocation lines and their interactions with other extended defects has not been captured yet by off-lattice KMC. This is an important challenge, since MD cannot simulate such kinetics at strain rates consistent with experiments. However, building an event catalog for a dislocation line – which contains thousands of interconnected possible sites where events may be take place – is a daunting task.

## 11    Conclusion

The last years have seen considerable developments with respect to accelerated atomistic methods, with access to ever more powerful computers and the introduction of new algorithms. These tools are giving modelers access to questions that could not even be asked in a recent past, increasing their interest for communities that are more turned to applications rather than methodological developments.

This additional interest will be beneficial to the field as it attracts new researchers with original ideas, knowledge, and skills, which will result in accelerated developments. Yet, as we have shown here, in spite of some limitations and questions, even in the current implementations, off-lattice kinetic Monte Carlo methods can deliver new insights for a wide range of problems dominated by activated diffusion.

The codes are available, and there is no reason today not to try them!

## 12    Code Availability

Various version of ART nouveau are available at http://normandmousseau.com. The kinetic ART package can be obtained freely by writing to MT or NM. The EON code is available at http://henkelmanlab.org/eon/.

## References

Alexander KC, Schuh CA (2016) Towards the reliable calculation of residence time for off-lattice kinetic Monte Carlo simulations. Model Simul Mater Sci Eng 24(6):65014. http://stacks.iop.org/0965-0393/24/i=6/a=065014?key=crossref.38e788234d74209ed2f8ad8b6b21fa51, https://doi.org/10.1088/0965-0393/24/6/065014
Althorpe S, Angulo G, Astumian RD, Beniwal V, Bolhuis PG, Brandão J, Ellis J, Fang W, Glowacki DR, Hammes-Schiffer S et al (2016) Application to large systems: general discussion. Faraday Discuss 195:671–698

Barkema GT, Mousseau N (1996) Event-based relaxation of continuous disordered systems. Phys Rev Lett 77(21):4358–4361

Béland LK, Mousseau N (2013) Long-time relaxation of ion-bombarded silicon studied with the kinetic activation-relaxation technique: microscopic description of slow aging in a disordered system. Phys Rev B 88(21):214201

Béland LK, Brommer P, El-Mellouhi F, Joly JF, Mousseau N (2011) Kinetic activation-relaxation technique. Phys Rev E 84(4):046704. https://doi.org/10.1103/PhysRevE.84.046704

Béland LK, Anahory Y, Smeets D, Guihard M, Brommer P, Joly JFF, Pothier JcC, Lewis LJ, Mousseau N, Schiettekatte F, Postale C, Centre-ville S (2013) Replenish and relax: explaining logarithmic annealing in ion-implanted *c*-Si. Phys Rev Lett 111(10):105502–105506. https://doi.org/10.1103/PhysRevLett.111.105502, http://arxiv.org/abs/1304.2991

Béland LK, Osetsky YN, Stoller RE, Xu H (2015a) Interstitial loop transformations in FeCr. J Alloys Compd 640:219–225

Béland LK, Osetsky YN, Stoller RE, Xu H (2015b) Slow relaxation of cascade-induced defects in Fe. Phys Rev B 91(5):054108

Béland LK, Samolyuk GD, Stoller RE (2016) Differences in the accumulation of ion-beam damage in Ni and NiFe explained by atomistic simulations. J Alloys Compd 662:415–420

Boulougouris GC, Frenkel D (2005) Monte Carlo sampling of a Markov web. J Chem Theory Comput 1:389–393

Boulougouris GC, Theodorou DN (2007) Dynamical integration of a Markovian web: a first passage time approach. J Chem Phys 127:084903

Brommer P, Béland LK, Joly JF, Mousseau N (2014) Understanding long-time vacancy aggregation in iron: a kinetic activation-relaxation technique study. Phys Rev B 90(13):134109–134117. https://doi.org/10.1103/PhysRevB.90.134109

Chill ST, Henkelman G (2014) Molecular dynamics saddle search adaptive kinetic Monte Carlo. J Chem Phys 140:214110

Chill ST, Stevenson J, Ruhle V, Shang C, Xiao P, Farrell J, Wales D, Henkelman G (2014a) Benchmarks for characterization of minima, transition states and pathways in atomic systems. J Chem Theory Comput 10:5476–5482

Chill ST, Welborn M, Terrell R, Zhang L, Berthet JC, Pedersen A, Jónsson H, Henkelman G (2014b) Eon: software for long time scale simulations of atomic scale systems. Model Simul Mater Sci Eng 22:055002

Duncan J, Harjunmaa A, Terrell R, Drautz R, Henkelman G, Rogal J (2016) Collective atomic displacements during complex phase boundary migration in solid-solid phase transformations. Phys Rev Lett 116(3):035701

El-Mellouhi F, Mousseau N, Lewis L (2008) Kinetic activation-relaxation technique: an off-lattice self-learning kinetic Monte Carlo algorithm. Phys Rev B 78(15):153202. https://doi.org/10.1103/PhysRevB.78.153202

Faken D, Jónsson H (1994) Systematic analysis of local atomic structure combined with 3D computer graphics. Comput Mater Sci 2:279–286

Fichthorn KA, Lin Y (2013) A local superbasin kinetic Monte Carlo method. J Chem Phys 138(16):164104

Guteŕrez M, Argaéz C, Jónsson H (2016) Improved minimum mode following method for finding first order saddle points. J Chem Theory Comput 13:125–134

Henkelman G, Jónsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 111:7010–7022

Henkelman G, Jónsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 113:9978–9985

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115(21):9657–9666. https://doi.org/10.1063/1.1415500

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113:9901–9904

Jay A, Raine M, Richard N, Mousseau N, Goiffon V, Hemeryck A, Magnan P (2017) Simulation of single particle displacement damage in silicon part II: generation and long time relaxation of damage structure. IEEE Trans Nucl Sci 64(1):141–148. https://doi.org/10.1109/TNS.2016.2628089, http://ieeexplore.ieee.org/document/7742370/

Joly JF, Béland LK, Brommer P, El-Mellouhi F, Mousseau N (2012) Optimization of the kinetic activation-relaxation technique, an off-lattice and self-learning kinetic Monte-Carlo method. J Phys Conf Ser 341:012007. https://doi.org/10.1088/1742-6596/341/1/012007, http://stacks.iop.org/1742-6596/341/i=1/a=012007?key=crossref.dfb01ebf3ff94111aa93a5794b3384f8

Joly JF, Béland LK, Brommer P, Mousseau N (2013) Contribution of vacancies to relaxation in amorphous materials: a kinetic activation-relaxation technique study. Phys Rev B 87(14):144204. http://link.aps.org/doi/10.1103/PhysRevB.87.144204

Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Berne BJ, Ciccotti G, Coker DF (eds) Classical and quantum dynamics in condensed phase simulations. World Scientific, Singapore, pp 385–404

Kim WK, Tadmor EB (2014) Entropically stabilized dislocations. Phys Rev Lett 112(10):105501

Koziatek P, Barrat JL, Derlet P, Rodney D (2013) Inverse Meyer-Neldel behavior for activated processes in model glasses. Phys Rev B 87:224105. https://doi.org/10.1103/PhysRevB.87.224105

Lu C, Jin K, Béland LK, Zhang F, Yang T, Qiao L, Zhang Y, Bei H, Christen HM, Stoller RE et al (2016) Direct observation of defect range and evolution in ion-irradiated single crystalline Ni and Ni binary alloys. Sci Rep 6:19994

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117:1–19

Machado-Charry E, Béland LK, Caliste D, Genovese L, Deutsch T, Mousseau N, Pochet P (2011) Optimized energy landscape exploration using the ab initio based activation-relaxation technique. J Chem Phys 135(3):034102. https://doi.org/10.1063/1.3609924, http://www.ncbi.nlm.nih.gov/pubmed/21786982

Mahmoud S, Trochet M, Restrepo OA, Mousseau N (2018) Study of point defects diffusion in nickel using kinetic activation-relaxation technique. Acta Mater 144:679–690. https://doi.org/10.1016/j.actamat.2017.11.021, http://www.sciencedirect.com/science/article/pii/S1359645417309643

Malek R, Mousseau N (2000) Dynamics of Lennard-Jones clusters: a characterization of the activation-relaxation technique. Phys Rev E 62(6):7723–7728. https://doi.org/10.1103/PhysRevE.62.7723

Marinica MC, Willaime F, Mousseau N (2011) Energy landscape of small clusters of self-interstitial dumbbells in iron. Phys Rev B 83(9):094119. https://doi.org/10.1103/PhysRevB.83.094119

Martínez E, Marian J, Kalos MH, Perlado JM (2008) Synchronous parallel kinetic Monte Carlo for continuum diffusion-reaction systems. J Comput Phys 227(8):3804–3823

McKay BD, Piperno A (2014) Practical graph isomorphism, II. J Symb Comput 60:94–112. https://doi.org/10.1016/j.jsc.2013.09.003, http://www.sciencedirect.com/science/article/pii/S0747717113001193

McKay BD et al (1981) Practical graph isomorphism. Congr Numer 30:45–87

Mousseau N, Barkema GT (1998b) Traveling through potential energy landscapes of disordered materials: the activation-relaxation technique. Phys Rev E 57:2419–2424. https://doi.org/10.1103/PhysRevE.57.2419

Munro LJ, Wales DJ (1999) Defect migration in crystalline silicon. Phys Rev B 59:3969

Nocedal J (1980) Updating quasi-Newton matrices with limited storage. Math Comput 35:773–782

Novotny MA (1995) Monte Carlo algorithms with absorbing Markov chains: fast local algorithms for slow dynamics. Phys Rev Lett 74:1–5

Novotny MA (2001) A tutorial on advanced dynamic monte carlo methods for systems with discrete state spaces. In: Stauffer D (ed) Annual reviews of computational physics IX. World Scientific, Singapore, pp 153–210

Ojifinni RA, Froemming NS, Gong J, Pan M, Kim TS, White J, Henkelman G, Mullins CB (2008) Water-enhanced low-temperature CO oxidation and isotope effects on atomic oxygen-covered Au (111). J Am Chem Soc 130(21):6801–6812

Osetsky YN, Béland LK, Stoller RE (2016) Specific features of defect and mass transport in concentrated FCC alloys. Acta Mater 115:364–371

Pedersen A, Jónsson H (2009) Simulations of hydrogen diffusion at grain boundaries in aluminum. Acta Mater 57:4036–4045

Pedersen A, Luiser M (2014) Bowl breakout: escaping the positive region when searching for saddle points. J Chem Phys 141(2):024109

Pedersen A, Henkelman G, Schiøtz J, Jónsson H (2009) Long time scale simulation of a grain boundary in copper. New J Phys 11:073034

Pedersen A, Berthet JC, Jónsson H (2012) Simulated annealing with coarse graining and distributed computing. Lect Notes Comput Sci 7134:34–44

Perez D, Luo SN, Voter AF, Germann TC (2013) Entropic stabilization of nanoscale voids in materials under tension. Phys Rev Lett 110(20):206001

Puchala B, Falk ML, Garikipati K (2010) An energy basin finding algorithm for kinetic Monte Carlo acceleration. J Chem Phys 132(13):134104. https://doi.org/10.1063/1.3369627, http://www.ncbi.nlm.nih.gov/pubmed/20387918

Raine M, Jay A, Richard N, Goiffon V, Girard S, Member S, Gaillardin M, Paillet P, Member S (2017) Simulation of single particle displacement damage in silicon part I: global approach and primary interaction simulation. IEEE Trans Nucl Sci 64(1):133–140. https://doi.org/10.1109/TNS.2016.2615133, http://ieeexplore.ieee.org/document/7582531/

Restrepo OA, Mousseau N, El-Mellouhi F, Bouhali O, Trochet M, Becquart CS (2016) Diffusion properties of Fe-C systems studied by using kinetic activation-relaxation technique. Comput Mater Sci 112:96–106. https://doi.org/10.1016/j.commatsci.2015.10.017, http://www.sciencedirect.com/science/article/pii/S0927025615006643, http://linkinghub.elsevier.com/retrieve/pii/S0927025615006643

Restrepo OA, Becquart CS, El-Mellouhi F, Bouhali O, Mousseau N (2017) Diffusion mechanisms of C in 100, 110 and 111 Fe surfaces studied using kinetic activation-relaxation technique. Acta Mater 136:303–314. https://doi.org/10.1016/j.actamat.2017.07.009, http://www.sciencedirect.com/science/article/pii/S135964541730558X

Shim Y, Amar JG (2005) Semirigorous synchronous sublattice algorithm for parallel kinetic Monte Carlo simulations of thin film growth. Phys Rev B 71:125432

Shim Y, Amar JG (2006) Hybrid asynchronous algorithm for parallel kinetic Monte Carlo simulations of thin film growth. J Comput Phys 212(1):305–317

Sinha AK (1972) Topologically close-packed structures of transition metal alloys. Prog Mat Sci 15:81

Sørensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112:9599–9606

Sørensen MR, Jacobsen KW, Jónsson H (1996) Thermal diffusion processes in metal-tip-surface interactions: contact formation and adatom mobility. Phys Rev Lett 77:5067–5070

Terentyev D, Malerba L, Klaver P, Olsson P (2008) Formation of stable sessile interstitial complexes in reactions between glissile dislocation loops in BCC Fe. J Nucl Mater 382(2):126–133

Terrell R, Welborn M, Chill ST, Henkelman G (2012) Database of atomistic reaction mechanisms with application to kinetic Monte Carlo. J Chem Phys 137:014105

Trochet M, Mousseau N (2017) Energy landscape and diffusion kinetics of lithiated silicon: a kinetic activation-relaxation technique study. Phys Rev B 96(13):134118. https://doi.org/10.1103/PhysRevB.96.134118

Trochet M, Béland LK, Joly JF, Brommer P, Mousseau N (2015) Diffusion of point defects in crystalline silicon using the kinetic activation-relaxation technique method. Phys Rev B 91(22):224106. https://doi.org/10.1103/PhysRevB.91.224106

Trochet M, Sauvé-Lacoursière A, Mousseau N (2017) Algorithmic developments of the kinetic activation-relaxation technique: accessing long-time kinetics of larger and more complex systems. J Chem Phys 147(15):152712. https://doi.org/10.1063/1.4995426

Trushin O, Karim A, Kara A, Rahman TS (2005) Self-learning kinetic Monte Carlo method: application to Cu(111). Phys Rev B 72(11):115401. https://doi.org/10.1103/PhysRevB.72.115401

Valiquette F, Mousseau N (2003) Energy landscape of relaxed amorphous silicon. Phys Rev B 68:125209. https://doi.org/10.1103/PhysRevB.68.125209

Vernon LJ (2010) Modelling the growth of $TiO_2$. Ph.D. thesis, Loughborough University

Vernon LJ (2012) PESTO: potential energy surface tools. https://github.com/louisvernon/pesto

Vernon L, Kenny SD, Smith R, Sanville E (2011) Growth mechanisms for $TiO_2$ at its rutile (110) surface. Phys Rev B 83(7):75412. https://doi.org/10.1103/PhysRevB.83.075412

Wales DJ (2002) Discrete path sampling. Mol Phys 100:3285–3305

Xiao P, Wu Q, Henkelman G (2014) Basin constrained $\kappa$-dimer method for saddle point finding. J Chem Phys 141:164111

Xu H, Osetsky YN, Stoller RE (2011) Simulating complex atomistic processes: on-the-fly kinetic Monte Carlo scheme with selective active volumes. Phys Rev B 84(13):132103. https://doi.org/10.1103/PhysRevB.84.132103

Xu H, Stoller RE, Osetsky YN, Terentyev D et al (2013) Solving the puzzle of <100> interstitial loop formation in BCC iron. Phys Rev Lett 110(26):265503

Xu H, Stoller RE, Béland LK, Osetsky YN (2015) Self-evolving atomistic kinetic Monte Carlo simulations of defects in materials. Comput Mater Sci 100:135–143

Xu L, Mei DH, Henkelman G (2009) Adaptive kinetic Monte Carlo simulation of methanol decomposition on Cu(100). J Chem Phys 131:244520

Zeng Y, Xiao P, Henkelman G (2014) Unification of algorithms for minimum mode optimization. J Chem Phys 140:044115

Zhou XW, Wadley HNG, Johnson RA, Larson DJ, Tabat N, Cerezo A, Petford-Long AK, Smith GDW, Clifton PH, Martens RL, Kelly TF (2001) Atomic scale structure of sputtered metal multilayers. Acta Mater 49:4005–4015

# Accelerated Molecular Dynamics Methods in a Massively Parallel World

R. J. Zamora, Danny Perez, E. Martinez, Blas Pedro Uberuaga, and Arthur F. Voter

## Contents

R. J. Zamora (✉)
Leadership Computing Facility, Argonne National Laboratory, Lemont, IL, USA
e-mail: rzamora@anl.gov

E. Martinez · B. P. Uberuaga
Materials Science and Technology Division, Los Alamos National Laboratory,
Los Alamos, NM, USA
e-mail: enriquem@lanl.gov; blas@lanl.gov

D. Perez · A. F. Voter
Theoretical Division T-1, Los Alamos National Laboratory,
Los Alamos, NM, USA
e-mail: danny_perez@lanl.gov; afv@lanl.gov

**Abstract**

As high-performance computing systems now rely on hardware parallelism for continuous improvements in performance, the timescales accessible to direct molecular dynamics (MD) remain limited by the relatively stagnant performance of a single processor. While spatial decomposition allows extremely large systems to be simulated for short periods of time, alternative methods are needed to extend these simulations to significantly longer timescales. Although parallel means to produce long-timescale trajectories have been known since the late 1990s, several newer methodologies falling under the umbrella of accelerated molecular dynamics (AMD) have recently been developed with parallelism in mind. In this chapter, we review the current state of the art in replica-based AMD and review several additional means of parallel scaling. Following a brief introduction to the serial AMD procedures, the chapter is organized into three general categories of parallel extensions: replication, speculation, and localization.

## 1 Introduction

Although CPU technology has generally kept pace with Moore's law over the past few decades, serial clock-speed performance has stagnated. Therefore, both hardware and software developers have been forced to rely on parallelism to continue improving the performance of both personal devices and leadership-scale machines. Within scientific computing, this has motivated significant efforts in the application and development of parallel algorithms, including many related to molecular dynamics (MD). For direct MD, these algorithms have focused on the most computationally intensive task: the calculation of forces on each atom for every sequential time-integration step. Increasing the simulation timescale has therefore been synonymous with decreasing the wall clock time (WCT) needed to calculate the forces (i.e., the WCT needed for each integration step).

For example, the so-called force-decomposition approaches employ shared memory techniques to speed up the calculation of forces. This approach is typically more scalable for computationally expensive potentials, like those based in quantum mechanics (Niklasson et al. 2016) or neural networks (Smith et al. 2017), where the amount of work is sufficiently large relative to the communication overhead. When the expensive potential can be approximated by a less-accurate (but faster) surrogate potential, the multi-time step parareal approach can be used to further leverage abundant parallel resources (Audouze et al. 2009). A significant speedup (with respect to conventional CPU architectures) can also be obtained using application-specific integrated circuits (ASIC) to optimize chip-level logic pipelines. Although this approach has been used to produce millisecond MD simulations for biomolecular systems (Shaw et al. 2008), the overall performance is still limited by the need for sequential time integration. Further scaling, i.e., across multiple nodes, usually relies on spatial decomposition methods. In fact, spatial decomposition methods have been so successful that modern petascale machines are routinely used to simulate billions

of atoms, or more, for short periods of time (Plimpton 1995; Germann and Kadau 2008; Zepeda-Ruiz et al. 2017).

While these approaches can be used for efficient weak scaling (i.e., efficiency can be maintained if the problem size increases in proportion to the number of cores), communication and synchronization overheads limit the potential for strong scaling (i.e., increasing the number of cores at fixed system size). This is because the true bottleneck of direct MD is the need to perform time-integration steps sequentially, requiring parallel algorithms to synchronize before (or after) each force calculation. As such, efforts to accelerate the force calculations alone are ultimately limited to the same timescales as serial MD (for small system sizes and fast interatomic potentials). Given that serial time integration is typically limited to microsecond timescales, conventionally parallel MD is still prevented from capturing critical long-time phenomena, such as thermally activated processes.

In this chapter, the typical material subjects of interest are solid-state configurations of atoms that are likely to evolve at finite temperature through so-called *rare events*. Here, a rare event corresponds to a state-to-state transition that is likely to occur on a timescale that is, on average, much larger than the time-integration step. States are defined as coarse regions of configuration space that are usually, but not exclusively, basins of attraction on the potential energy surface (PES). Good state definitions are such that a typical MD trajectory spends a long time within a state before moving on to another one. When the evolution of a system is truly dominated by rare events, a direct MD simulation becomes unlikely to pass through more than a small handful of states.

To capture dynamical processes that are driven by rare events, specialized simulation techniques are typically needed to bypass the limitations of sequential time integration. This chapter focuses specifically on the accelerated molecular dynamics (AMD) family of techniques. In a nutshell, the AMD strategy to long-timescale simulations is to leverage the disparity between rapid thermal fluctuations within a state, and activate state-to-state transitions, to controllably accelerate the simulated dynamics. AMD is distinct from adaptive kinetic Monte Carlo (AKMC) (Henkelman and Jónsson 2001), because MD is used to discover every transition in the simulated trajectory, avoiding the need to build a complete rate catalog for every visited state.

Beginning in 1997, three AMD methods were proposed to accelerate the generation of state-to-state trajectories for rare-event systems, namely, hyperdynamics, which relies on a direct manipulation of the underlying PES (Voter 1997), parallel replica dynamics (PRD), which relies on the parallel accumulation of MD time (Voter 1998), and temperature-accelerated dynamics (TAD), which relies on the exponential increase in transition rates as the temperature is increased (Sørensen and Voter 2000). For the sake of a self-contained review, we will briefly reintroduce these *original* AMD methods and refer the reader to the first issue of the *Handbook of Materials Modeling* for a more in-depth discussion of the theoretical details (Uberuaga et al. 2005). In this chapter, our primary goal is to highlight recent and ongoing efforts to improve upon the performance of established techniques, using parallel algorithms.

Of the original AMD methodologies, only PRD is a fundamentally parallel approach. However, recent work has shown that a number of parallel techniques can also be leveraged by the other methods. In this chapter, we will focus our discussion on the most recent advances in this spirit. Generally, these approaches fall into (at least) one of the following three categories: replication, speculation, and localization.

## 2   Serial AMD

Before discussing recent developments, we first review the methodologies that do not require parallel computing resources to achieve a speedup (boost) with respect to direct MD. We present only the fundamental concepts of these methodologies to provide a foundation for the following discussions on replication, speculation, and localization. The reader is referred to the referenced resources for a more comprehensive discussion.

## 2.1   Hyperdynamics

As in all the AMD methods, the goal of hyperdynamics is to use a modified MD-based approach to find an appropriate first-escape path from the current state more quickly than MD would (Voter 1997). Here, "appropriate" means that the relative probability of choosing a particular escape path is the same as it would be in direct MD (i.e., proportional to its rate constant) and that the time at which the escape occurs is drawn from the same distribution.

Hyperdynamics achieves this by executing dynamics on a modified potential surface, $V(X) + V^b(X)$, in which the nonnegative bias potential $V^b(X)$ has been added to the original potential to partially "fill in" the basin. The main requirement on $V^b(X)$ is that it be zero at all dividing surfaces between states. The derivation assumes that transition state theory (TST) holds on both the original and modified potentials; i.e., there are no correlated events such as recrossings or multiple hops, so every crossing of a dividing surface is followed by loss of memory in the new state. In this limit, the transition rate between states is directly expressible as a thermodynamic quantity: the equilibrium flux of trajectories through the dividing surface between states. When this is a good approximation, it can be shown that hyperdynamics will increase all transition rates out of a given states by the same factor, the so-called boost factor, which is defined as:

$$\eta^{hyper} = \langle \exp(+V^b(X(t))/k_B T) \rangle. \tag{1}$$

In other words, the state-to-state dynamics on $V(X)$ and $V(X) + V^b(X)$ are equivalent up to a rescaling of time by $\eta^{hyper}$ to account for the fact that physical time flows $\eta^{hyper}$-times faster on the biased potential as compared to the original one.

The design of an effective, and valid, bias potential is a major project in itself, as it must be zero at all dividing surfaces in spite of having no explicit knowledge of where those surfaces may be. While a number of alternatives are available, Miron and Fichthorn (2003) proposed a particularly effective, geometry-based form, expressed in terms of "bond" lengths, where a bond is defined as any pair of atoms within a threshold distance of each other (typically the threshold would be set a little larger than a nearest-neighbor distance in the native crystal form). The idea behind the bond boost bias is to make $V^b = 0$ for any geometry in which one or more bonds is distorted (either shortened or lengthened) beyond a threshold value (e.g., 30%). For bond $j$, the "equilibrium" bond length $r_j^{eq}$ is the distance between the pair of atoms when the system potential energy has been minimized (i.e., at the basin-bottom configuration $R_{min}$), and the relative distortion of bond $j$ is defined as

$$\varepsilon_i = [r_j - r_j^{eq}]/r_j^{eq}. \tag{2}$$

In the "simple bond boost method" (Perez et al. 2009) (SBB), the bias potential depends only on the distortion $\varepsilon_{max}$ of the most distorted bond in the system,

$$V_{SBB}^b = V_{SBB}^b(\varepsilon_{max}) = \begin{cases} V_{max}^b \left[1 - (\varepsilon_{max}/q)^2\right], & |\varepsilon_{max}| < q \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $q$ is the threshold distortion for complete shutdown of the bias and $V_{max}^b$ is the maximum possible bias energy. (Note that the original bond boost method requires the use of an envelope function to allow multiple bonds to contribute to the total bias potential.)

## 2.2 Temperature-Accelerated Dynamics (TAD)

Temperature-accelerated dynamics efficiently predicts the escape path from a given state by performing basin-constrained MD (BCMD) at an elevated temperature compared to the actual physical system of interest (Sørensen and Voter 2000; Zamora et al. 2016a). In BCMD, high-temperature MD is performed as usual, but whenever a transition is detected, the system is placed back into the original state (effectively constraining the dynamics to a single PES basin). This is repeated until one can assess that the transition that should have occurred first at the original temperature has been observed at the high temperature. To project the observed transition times at high temperature onto a hypothetical low-temperature time line, the method assumes that harmonic transition state theory (HTST) holds (Vineyard 1957). Using this approximation, an observed high-temperature transition along pathway $j$ corresponds to a hypothetical low-temperature transition at time

$$t_j^{Low} = t_j^{High} \exp\left[E_j \left(\frac{1}{k_B T_{Low}} - \frac{1}{k_B T_{High}}\right)\right]. \tag{4}$$

Here, $k_B$ is the Boltzmann constant, $T_{\text{High}}$ is the high temperature used for BCMD, $T_{\text{Low}}$ is the low temperature of interest, and $t_j^{\text{High}}$ is the observed transition time at $T_{\text{High}}$. The value of the energy barrier for pathway $j$, $E_j$, is typically calculated using the nudged elastic band (NEB) method of Henkelman et al. (2000).

Since an elevated temperature tends to favor high-barrier transition pathways, it is important that BCMD be performed until it becomes sufficiently unlikely that a later transition at $T_{\text{High}}$ will correspond to the earliest transition at $T_{\text{Low}}$. In practice, this "stop" time, $t_{\text{Stop}}^{\text{High}}$, is defined in terms of the shortest low-temperature transition time, $t_{\text{Short}}^{\text{Low}}$, an adjustable uncertainty parameter, and an assumed minimum pre-exponential factor.

For the following discussion, we define $t_{\text{Winner}}^{\text{High}}$ as the value of $t_j^{\text{High}}$ corresponding to $t_{\text{Short}}^{\text{Low}}$ when the TAD procedure is completed for the current state. In order to predict a long-time trajectory involving many sequential state-to-state transitions, this procedure is typically repeated after moving the system into the state corresponding to the winning high-temperature transition time, $t_{\text{Winner}}^{\text{High}}$, once $t_{\text{Stop}}^{\text{High}}$ has been reached.

### 2.2.1 Synthetic-Mode TAD

In practice, the performance of any AMD method is limited by the rate of state-to-state transitions in the simulated physical system, i.e., the boost is typically controlled by the fastest processes that change the state of the system. For users and developers of rare-event simulation techniques, this is often referred to as the *low-barrier* problem. When a system is able to visit a large number of distinct states without having to overcome significant energy barriers (relative to $k_B T$), the standard TAD procedure will have very little opportunity to provide a boost. For many realistic systems, local clusters of neighboring states will be connected by energy barriers that are much lower than those separating the cluster from external states. This type of cluster is often referred to as a superbasin of states, because it can be defined as a basin of other PES basins. Since the system can typically move among the internal states of a superbasin at a much faster rate than it can escape into an external state, an MD trajectory is likely to revisit the internal states many times before making its way to another (super) basin.

One significant advantage of TAD is that its theoretical foundation allows a so-called *synthetic mode* to be used whenever superbasins are encountered (Sørensen and Voter 2000). When individual states are revisited many times, synthetic mode provides a means to leverage the total high-temperature BCMD time that has already been accumulated over all previous visits. The basic idea is that $t^{\text{High}}$ need not be reset to zero when the TAD procedure is initiated in a state that has already been explored. Instead, the high-temperature clock can pick up where it left off during the previous visit, as long as every previously observed escape pathway is explicitly represented as a future transition at a predetermined time, and the projected low-temperature time is offset from the previous visit.

The predetermined transition times for a given escape pathway can simply correspond to previously observed high-temperature transitions that were not accepted. However, if the escape path has a known (or approximately known) transition rate, a corresponding escape time can be randomly sampled. If the low-temperature transition rate can be accurately estimated for a given pathway, using either observed statistics or a direct calculation, the pathway can be promoted to *synthetic mode*. Once a pathway is in synthetic mode, all future transition times correspond to synthetic predictions that are determined using a sampling procedure similar to kinetic Monte Carlo (KMC). This means that subsequently observed MD transitions along synthetic pathways no longer need to be projected onto the low-temperature time line.

It is clearly advantageous to allow $t^{High}$ to grow as large as possible within a single state, because the fixed increment in $t^{Low}$ projected by each MD step grows exponentially with $t^{High}$. After a state has been revisited many times, the synthetic-mode procedure can result in the accumulation of enough $t^{High}$ to accept many synthetic transitions at once. After this has happened in every internal state of a superbasin, the performance of synthetic-mode TAD can approach that of KMC, until the system escapes from the superbasin. In practice, the synthetic mode is invaluable for the simulation of realistic material defects. For this reason, the procedure will also play a critical role in the performance of some parallel TAD extensions discussed below.

Now that hyperdynamics and TAD have been briefly introduced, the stage has been set to describe their parallel extensions. But, before doing so, we will first describe the basic procedure behind the remaining original AMD method, PRD, and introduce the broader concept of parallel replication.

## 3 Replication

Although PRD was first introduced in 1998 (Voter 1998), the method has recently experienced a coming of age, both due to the increasing availability of parallel computing resources and to a better mathematical understanding of the underlying theory (Le Bris et al. 2012). Also key to its usefulness is the fact that the general time-wise decomposition approach used by PRD, referred to here as *parallel replication*, can be extended to improve the performance of other methodologies. In the following, we will review the current state of the original PRD approach and present recent PRD+X efforts to combine parallel replication with other AMD approaches. In order to best describe the fundamentals of parallel replication, we will start by reviewing the basics of PRD. For a more comprehensive introduction, the reader is referred to the recent review by Perez et al. (2015).

### 3.1 Parallel Replica Dynamics (PRD)

For PRD, we assume that there are $N_r$ dedicated computing units available, called *replica* processes, that are each capable of carrying out independent MD

simulations. Starting with the system in some initial state, the PRD algorithm follows the steps illustrated in Fig. 1:

1. Let one replica carry out MD until it remains in the same state for a correlation time ($\tau_c$). This is called the decorrelation stage (blue).
2. Broadcast the decorrelated state to the other $N_r - 1$ replicas (orange). These replicas then independently carry out MD until they spend $\tau_c$ in the initial decorrelated state, rejecting transitions whenever they are detected. This is called the dephasing stage (red), because it is used to ensure that each replica is statistically independent.
3. After successfully dephasing/decorrelating, each replica enters the parallel stage (green), where it independently carries out MD, periodically checking whether a transition occurred (note that different replicas can enter the parallel stage at different times).
4. Once a transition is detected, the replica on which the transition occurred triggers a new decorrelation step, initiating a new algorithm cycle. At each cycle, the total simulation time is incremented by the duration of the decorrelation stage, plus the sum of the MD times accumulated on all replicas during the parallel stage (blue + green).

This algorithm has been shown to produce an arbitrarily accurate state-to-state dynamics, simply by adjusting the value of $\tau_c$ (Le Bris et al. 2012). For a physical system with a well-defined global transition rate of $k_g$, the speedup of PRD over direct MD can be written as

$$\eta^{\mathrm{PRD}} = N_r \left( (1 + f_Q)(1 + 2N_r k_g \tau_c) \right)^{-1},$$  (5)

where $f_Q$ is the fractional number of additional force calculations, per MD step, needed to periodically assess whether a transition has occurred. Note that Eq. (5) assumes that serial MD is performed by each replica process.
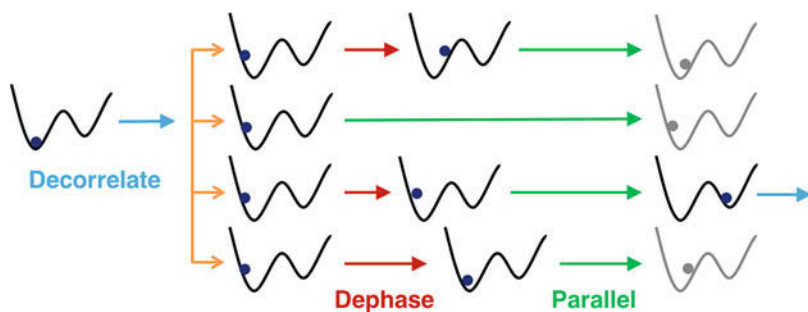


**Fig. 1** Illustration of the PRD algorithm for a simple energy landscape where the states are defined as local basins of attraction on the PES (black curves). For each idealized geometry, the circle indicates the position of the system on the landscape. The different stages of the algorithm are indicated by colors: decorrelation (blue), broadcast (orange), dephasing (red), and parallel (green)

This analysis suggests that the parallel efficiency of PRD is inversely proportional to $k_g$ and that perfect speedup of $N_r$ can be obtained only in the limit where $k_g \rightarrow 0$. This general rule is not specific to the performance of PRD but also describes the upper bound on the other original AMD methods. That is, while both standard hyperdynamics and TAD can provide a significant boost factor without the need for parallel computing resources, the optimal boost should be comparable to PRD with $N_r \rightarrow \infty$. When combining replication-based parallelism with these other AMD approaches (PRD+X), as we will discuss below, parallelism alone cannot move the performance of these methods beyond the theoretical limits of pure PRD on a sufficiently large number of cores. However, if such a number of cores are not available, a PRD+X combination might allow for the optimal boost to be achieved with more accuracy than X-only, since it will allow for a less aggressive bias potential in hyperdynamics or a lower high temperature in TAD.

To help express the advantages and limitations of replication, as well as the other parallel methodologies to be discussed below, it is useful to directly compare the available performance of PRD with that of spatially parallel MD. Following the analysis of Martínez et al. (2014), Fig. 2 shows a comparison of the available simulation timescale (x-axis) and length scale (y-axis) for both MD and PRD. The plot assumes a 24-hour simulation using a classical interatomic potential, with an escape rate per atom of $k = 10^5$ (s atom)$^{-1}$, a 1 fs MD time step, an execution
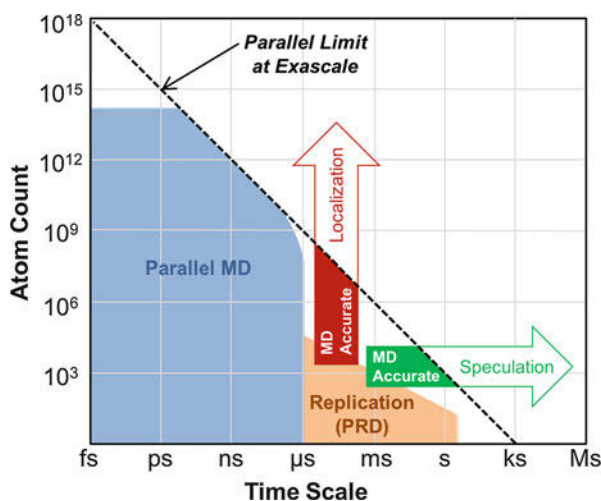


**Fig. 2** Comparison of the simulated timescales and length scales accessible to both MD and AMD, given a 24-hour run time on a hypothetical exascale machine. The specific shapes of the MD and PRD *polygons* are based on benchmark simulations and analysis presented in reference (Martínez et al. 2014). The shape of the PRD region assumes a per-atom escape rate of $k = 10^5$ (s atom)$^{-1}$, a 1 fs MD time step, a processor speed of 1 μs per atom per time step, and 12 million processors on the whole machine. The theoretical parallel limit is shown as a dashed black line. The intended direction of improved performance is also shown for speculation and localization

time of 1 μs per atom per time step on a single processor, and 12 million processors on the whole machine. Note that the size-dependent escape rate is here given by $k_g = k N_{at}$, where $N_{at}$ is the number of atoms in the system.

Since PRD must rely on parallelism to access extended timescales, limits on available computing resources will always lead to an upper limit on performance. In Fig. 2, this theoretical limit is depicted by a dashed black line that divides the plot into upper and lower triangles. The x-intercept of this line can be defined by the $k_g \rightarrow 0$ limit of Eq. (5). It is important to recognize that, unlike PRD, neither hyperdynamics nor TAD are strictly prevented from accessing the upper triangle of this performance landscape, because they do not rely on parallelism to achieve their boost. However, at increasingly large computational scales, the actual performance of AMD methods is increasingly more likely to depend on the value of $k_g$ than on the parallel limit.

Figure 2 clearly shows that PRD can provide access to a significant portion of spatiotemporal simulation space that is inaccessible to direct MD at the exascale. However, the plot also illustrates that the global transition rate can prevent PRD from reaching the theoretical parallel limit on a large machine. Breaking this limitation is the motivation behind many of the novel speculation and localization approaches that will be described below.

## 3.2    PRD+X

### 3.2.1    Parallel Replica Hyperdynamics (PRH)

The parallel replica hyperdynamics method (PRH), corresponding to the combination of PRD and hyperdynamics, is almost as old as its two components (Voter and Germann 1998). Given an appropriate hyperdynamics bias potential, combining these two methods is straightforward: PRD is performed as usual, except the dynamics on each replica is performed on the biased potential surface, $V(X) + V^b(X)$, to further accelerate the local transition rate. Since each replica process will experience a boosted transition rate of $k_g \eta_r^H$, where $\eta_r^H$ is the hyperdynamics boost factor obtained on each replica, the new algorithm can be expected to provide a speedup of

$$\eta^{PRH} = N_r \eta_r^H \left( (1 + f_Q)(1 + 2 N_r \eta_r^H k_g \tau_c) \right)^{-1}. \tag{6}$$

This approximate equation demonstrates that the available boost factors of PRD and hyperdynamics do not simply multiply, because the efficiency of PRD will be inhibited by the boosted transition rate experienced by each replica. The maximum possible speedup in PRH therefore tends to that of PRD as the number of processors increases. However, the hybrid technique can be very valuable when resources are limited. For example, in cases where $\eta^{PRD}$ is limited by $N_r$, rather than $k_g$, it is very likely that PRH will outperform PRD-alone.

### 3.2.2   Parallel Replica TAD (ReplicaTAD)

The newest PRD+X combination is the parallel replica TAD method (ReplicaTAD) (Zamora et al. 2018). ReplicaTAD uses one computing unit to carry out the original TAD procedure, including all NEB calculations, and $N_r$ to carry out high-temperature MD using PRD.

In contrast to original PRD however, ReplicaTAD uses replication to reach $t_{\text{Stop}}^{\text{High}}$ as quickly as possible, rather than to decrease the waiting time to observe a single transition (although this is also achieved). Since TAD uses BCMD rather than direct MD, the parallel replication procedure can be performed in a way that is more efficient than the original PRD. For example, individual replicas must dephase only when they locally observe a transition, and, when this occurs, the final state does not have to be broadcasted to every other replica but only to the master.

Like PRH, ReplicaTAD is powerful but unlikely to outperform pure PRD if unlimited parallel resources are available (this does not consider the case where the synthetic mode can be used by ReplicaTAD to escape superbasins). Although BCMD allows for the use of a more efficient replication procedure, the elevated temperature means that the parallel replica efficiency is worse than it would have been at the low temperature of interest. However, the real advantage of ReplicaTAD is that computational resources can be used to improve the accuracy of the TAD procedure. Since the kinetics in most material systems departs from HTST predictions when the temperature becomes too large, TAD is most accurate when $T_{\text{High}}$ approaches $T_{\text{Low}}$. While a low $T_{\text{High}}$ setting is likely to produce a meager boost in serial TAD, adding parallel replication can produce a boost that is equivalent to, if not larger than, the serial TAD boost with an optimal $T_{\text{High}}$ setting.

## 4   Speculation

Speculative execution is a well-established technique used across various applications such as distributed computing and modern CPU logic. For many modern pipe-lined CPU architectures, for example, the use of branch prediction is often required to achieve optimal performance. The basic idea is that, when excess computing resources are available, it may be preferable to perform certain tasks before there is any guarantee that the task is necessary, so long as the early execution of said task has the potential to decrease the overall execution time.

Within the context of AMD, parallel speculation generally refers to the parallel scaling of a parent method by leveraging speculative execution in states that the physical system is expected to visit (or revisit) in the future. Execution becomes speculative in nature when the simulated trajectory has yet to require additional MD in that state to continue. Although it is generally not straightforward to make accurate predictions of future needs, there are at least two situations in AMD where this can be done relatively simply. The first of these cases occurs during any TAD simulation. The second case occurs when the trajectory is trapped in a superbasin. Next, we discuss two recent approaches intended to address these speculative execution opportunities.

## 4.1    Speculation in TAD

The opportunity for efficient speculation within the TAD method arises from the use of BCMD. While transitions are explicitly rejected before the proper low-temperature transition is finally identified, their occurrence can be used to accurately *predict* the next official state along the trajectory (note that we use the term *official state* to describe a state that is included in the final state-to-state trajectory).

### 4.1.1    Speculatively Parallel TAD (SpecTAD)

The speculatively parallel TAD (SpecTAD) method uses speculative execution to scale the serial TAD algorithm across multiple processors (Zamora et al. 2016a). As mentioned above, the nature of BCMD leads to an ideal opportunity for speculative execution. After a transition pathway is first observed at high temperature, an idle processor can be used to immediately begin executing the TAD procedure in the final state if it is possible that the transition will ultimately be accepted. For many real TAD simulations, speculative execution can allow significant parallel scaling, because $t_{\text{Stop}}^{\text{High}}$ can be more than an order of magnitude longer than $t_{\text{Winner}}^{\text{High}}$. Since Eq. (4) is used to update $t_{\text{Short}}^{\text{Low}}$ as soon as the NEB is performed for each high-temperature transition, it is also straightforward to decide which speculative states should be prioritized for exploration on the available CPU resources.

The basic SpecTAD procedure requires relatively simple modifications to the traditional TAD approach. The most significant change is that a new master process is needed to build the long-time trajectory of many sequential state-to-state transitions. This master process does not need to perform any MD or know much about the TAD procedure. Instead, this process simply manages some number of speculation processes that are each responsible for performing a modified TAD procedure in a single assigned state. When a speculation process is not performing TAD, it simply waits for the master process to assign it a new state (in practice, this idle waiting time is what typically limits the parallel scaling of SpecTAD). As illustrated in Fig. 3, the simulation will begin with the master sending the initial state geometry (State A) to a random process (Process 0). Note that the processes in Fig. 3 are labeled in the order that they are first assigned a state but that any idle process could be used to explore each state. When a process is assigned a new state, it executes the TAD procedure with the following modifications:

1. When $t_{\text{Short}}^{\text{Low}}$ is replaced by a new transition at $T_{\text{High}}$, using Eq. (4), a speculation message must be sent to the master process. This message will include the geometry of the neighboring state corresponding to the new $t_{\text{Short}}^{\text{Low}}$.
2. When $t_{\text{Stop}}^{\text{High}}$ is reached, the master process must be alerted that the most recent speculation from this process actually corresponds to the winning transition for the corresponding state visit and that this process is now idle.
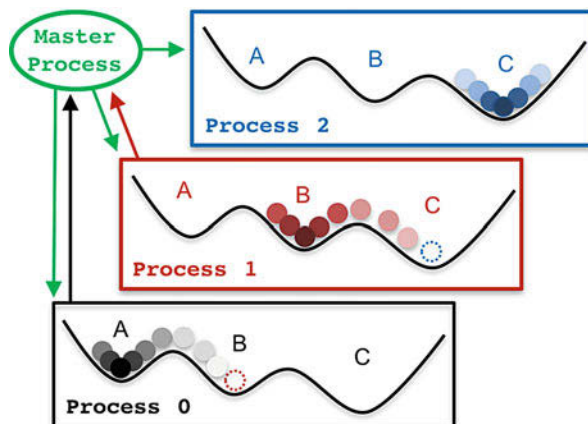
**Fig. 3** Schematic illustration of SpecTAD on an idealized 2-D potential energy landscape. Each rectangle corresponds to a distinct speculation process, which executes a modified TAD procedure in a single assigned state. The simulation begins with the master process assigning State A to Process 0, where the red-dashed line circle corresponds to a detected transition, on Process 0, to State B. Once informed of the transition, the master process assigns State B to Process 1, and the same process is repeated when a transition is detected in State B (blue-dashed line circle)

3. Processes must occasionally probe for a kill message from the master. This message indicates that this process is currently exploring a state that is no longer a priority.

Although the schematic illustration in Fig. 3 excludes the killing of speculation processes to reassign work, the intended execution flow is clearly shown: as soon as Process 0 detects a transition that allows it to set (or reset) $t_{\text{Short}}^{\text{Low}}$, the master process will assign the next expected state (State B) to an idle speculation process (Process 1). This same procedure is then repeated in State B, leading to a third speculation process (Process 2) exploring State C. Since $t_{\text{Stop}}^{\text{High}}$ can be orders of magnitude longer than $t_{\text{Winner}}^{\text{High}}$ for each distinct speculation process, all three processes can perform much of their work in parallel.

In contrast to replication-based AMD methods like ReplicaTAD, the parallel efficiency of SpecTAD is clearly dependent on the total number of transitions that occur in the official low-temperature trajectory. Since the only source of parallel speedup is the concurrent execution of TAD in speculative future states, the simulation of a trajectory with N transitions could never use more than N cores efficiently. In practice, this is not a stringent limitation, as long-time simulations often include thousands of transitions or more. Instead, the maximum parallel speedup is more typically limited by the ratio: $\tilde{w}_{\text{Stop}}/\tilde{w}_{\text{Winner}}$, where $\tilde{w}_{\text{Stop}}$ corresponds to the average wall-clock time (WCT) needed to reach the stop time for a given state and $\tilde{w}_{\text{Winner}}$ corresponds to the average WCT needed to find the winning transition. Since each speculation process must discover its winning transition before subsequent states

can be explored, this ratio gives an approximate upper bound on the attainable parallel speedup.

### 4.1.2 Synthetic-Mode SpecTAD

Like the original TAD method, SpecTAD is particularly well-suited for the simulation of physical systems containing superbasins of states. This is because each speculation process can employ synthetic-mode transitions to leverage previously accumulated $t^{\text{High}}$ in revisited states. More importantly, however, is the fact that the application of synthetic transitions can actually improve the overall parallel efficiency of SpecTAD. Although the synthetic-mode procedure is likely to reduce the value of $\tilde{w}_{\text{Stop}}$, the procedure will even more dramatically reduce the value of $\tilde{w}_{\text{Winner}}$. Since the procedure requires that future time stamps be assigned to all escape pathways observed during previous visits, it becomes likely that the winning transition out of a revisited state can be accurately predicted before any additional MD is performed. Therefore, once a SpecTAD simulation becomes stuck within a deep superbasin of known states, it becomes possible for the simulation to use all available resource to escape from the superbasin.

### 4.1.3 Hybrid SpecTAD

Although SpecTAD is certainly a powerful extension of the original TAD method, in general, pure speculation is unlikely to be the most efficient use of parallel resources. For example, when a physical system exhibits anharmonic behavior above some modest temperature, an accurate simulation may require that the value of $T_{\text{High}}$ be set relatively close to $T_{\text{Low}}$. In such cases, ReplicaTAD will often outperform SpecTAD, because the $\tilde{w}_{\text{Stop}}/\tilde{w}_{\text{Winner}}$ ratio becomes too small for pure speculation to scale efficiently. However, in that case, a flexible combination of both speculation and replication can be used to efficiently improve the TAD boost for any $T_{\text{High}} > T_{\text{Low}}$.

The basic idea behind the *hybrid* SpecTAD method (Zamora et al. 2018) is to allow for the combination of both speculation and replication. This is accomplished by modifying the SpecTAD algorithm to utilize $N_r$ local replica processes within each of the $N_s$ speculation processes. Using this scheme, which is schematically illustrated in Fig. 4, each speculation process performs a ReplicaTAD procedure, using $N_r + 1$ computing units. It is useful to note that this hybrid method will improve upon the performance of conventional SpecTAD, even when $N_r = 1$, as NEB calculations and MD can then be executed concurrently.

In practice, the most important advantage of hybrid SpecTAD is that it provides a straightforward way to trade accuracy and resources. It becomes practical to start with an MD-accurate method with $T_{\text{High}} = T_{\text{Low}}$ and then to raise the high temperature until the required timescale can be accurately accessed, using whatever computing resources are available. The application of an elevated temperature also opens up the opportunity to promote repetitive events into synthetic mode – possibly leading to KMC-like performance.

The development of this new hybrid method demonstrates the advantages of combining multiple forms of parallelism. In contrast to MD-accurate methods,
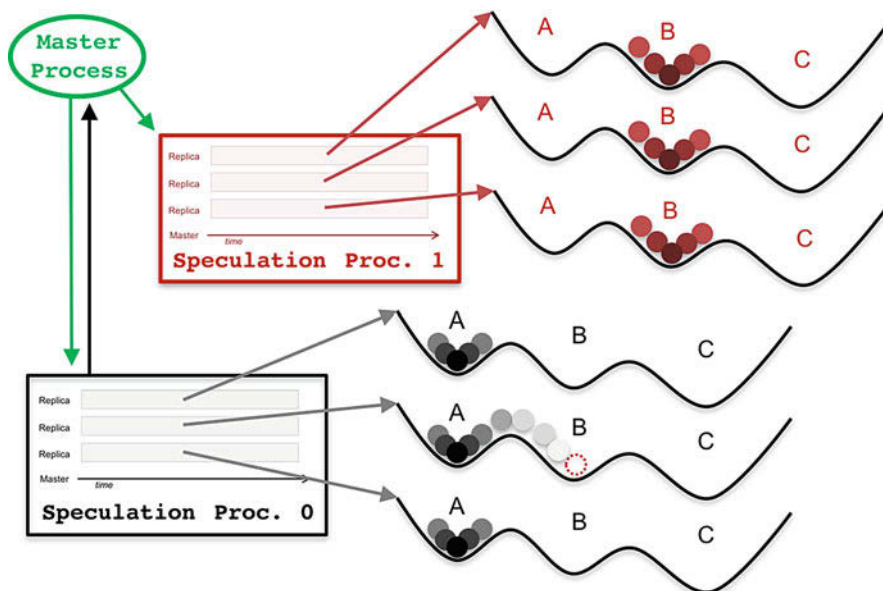
**Fig. 4** Schematic illustration of hybrid SpecTAD on an idealized 2-D potential energy landscape. Each rectangle corresponds to a distinct speculation process, which uses a local master process, and $N_r$ local replica processes, to execute a modified TAD procedure in a single assigned state. Like conventional SpecTAD, a global master process is used to manage the assignments of states to available speculation processes

however, TAD-based algorithms are not limited to the lower triangle presented in Fig. 2. In practice, the boost available from hybrid SpecTAD can be many orders of magnitude larger than the theoretical parallel limit. For example, recent work by Zamora et al. (2016b) has leveraged both synthetic-mode transitions and speculation to achieve boost factors greater than $10^{15}$ over MD, corresponding to years of simulated time. By combining replication and speculation, extreme boost factors like this should become even more accessible, because parallel resources are more efficiently used to accelerate the rate at which internal superbasin states are explored.

## 4.2    Speculation in PRD

The TAD formalism offers a straightforward way to speculate which transitions are likely to occur in the future, as a transition is always observed at high temperature before it is officially accepted as a valid low-temperature transition. This delay opens a window of opportunity to overlap calculations in different states. Such a delay is however not available in PRD, since transitions are accepted as soon as they are observed. Nonetheless, if one can hazard a prediction that some states might be visited, or revisited, in the future, it can be leveraged to improve efficiency of replica methods like PRD.

### 4.2.1 Parallel Trajectory Splicing (ParSplice)

The parallel trajectory splicing (ParSplice) method (Perez et al. 2016) is an extension of PRD, designed to dramatically improve the parallel efficiency when states are likely to be revisited. Like PRD, the ParSplice boost is obtained through replica-based parallelism. However, the worker processes are no longer synchronized in the same states. Indeed, the key idea in ParSplice is to factor the problem of generating a long state-to-state trajectory into independent tasks that can be carried out concurrently, independently, and asynchronously. These tasks are to generate so-called trajectory *segments* such that the system remained for at least $\tau_c$ in a given state prior to the beginning of the segment and at least $\tau_c$ in a given state (potentially different from one where the segment began) prior to the end of the segment. It can be shown using the same formalism developed for PRD (Le Bris et al. 2012) that such segments can be assembled (or spliced) into a proper state-to-state trajectory as long as the new segment begins in the same state where the trajectory currently ends.

The improvement in efficiency over PRD is primarily due to aggressive book-keeping: first, segments that cannot be immediately spliced into the main state-to-state trajectory are stored in a database for future use in case the trajectory revisits their starting state at a later time; second, end points of segments, which are by construction proper starting points for subsequent segments, are stored and reused in order to avoid having to re-dephase in states that have been previously visited. Taken together, this insures that computer cycles invested in MD can potentially be leveraged to create the state-to-state trajectory.

Efficiency is further improved through speculative segment execution. Indeed, as shown in Fig. 5, ParSplice workers are not constrained to run in the same state. Instead, segments can be simultaneously generated in multiple states according to the *expected likelihood* that they will be spliced into the trajectory in the future. Operationally, this likelihood optimization is carried out using simulations of the possible future evolution of the trajectory based on a kinetic model that is constructed on the fly. As with SpecTAD, this procedure is speculative in nature because one cannot certify that the segment that is about to be generated will be indeed spliced into the trajectory. Exploiting this uncertainty however enables a considerable increase in the parallelism available to the algorithm – one can parallelize over the possible futures of the trajectory, in addition to its present state – thereby dramatically increasing scalability in some cases.

ParSplice performs best in the rather common situation where the trajectory is trapped in a superbasin of states (Perez et al. 2018). In this case, the same states will be revisited over and over, which maximizes the probability that completed segments are already available at the current end of the trajectory and that dephased starting points are available to efficiently generate additional segments. Further, the evolution of a trajectory in a superbasin is statistically predictable. This allows for the optimal assignment of additional segments in the different states within the superbasin. It was recently shown that ParSplice can maintain excellent performance as long as the escape out of the superbasin is slow, even when individual intra-basin transitions occur so rapidly that a synchronous PRD approach would be inefficient
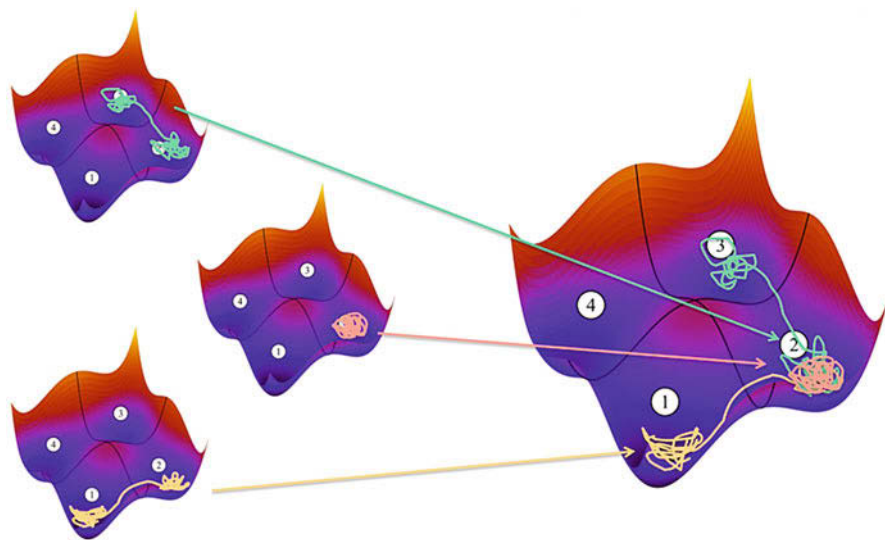
**Fig. 5** Illustration of the ParSplice method. Trajectory segments are independently and concurrently generated in multiple states. These segments are then assembled into a long state-to-state trajectory by splicing together segments whose end points match

(Perez et al. 2018). This enables one to leverage massive parallel resources to investigate complex systems where the distribution of transition rates is very wide.

It is useful to recognize that ParSplice is actually a natural combination of both parallel replication and speculation. For this reason, the general resource-management approach can be used to efficiently scale the performance of hyper-dynamics or TAD on parallel resources. In the case of TAD, there is no reason that a master process needs to directly splice the generated segments into a single official trajectory. Instead, the master can separately splice many distinct high-temperature trajectories, corresponding to both current and speculative states, and then use Eq. (4) to project each of these spliced trajectories onto a single (official) low-temperature trajectory. When compared to the existing version of hybrid SpecTAD, the incorporation of a ParSplice-based resource management procedure would effectively produce an adaptive load balancing mechanism to optimally choose the number of replication and speculation processes in use. Given the significant value of efficiency in the world of high-performance computing, this expected outcome is a powerful motivation for further development.

## 5    Localization

While replica and speculation-based parallelism can allow an AMD method to access significantly longer timescales than direct MD, other techniques are often needed to allow the boost to be maintained at larger length scales. Indeed, the boost

of a traditional AMD method often scales poorly with an increase in the system size, because the global transition rate will also proportionally grow. For example, an infinite system will always experience at least one transition occurring at some location at any given time. Since PRD, TAD, and hyperdynamics all rely on a separation of timescales between thermal fluctuations and state-to-state transitions, their performance suffers when these timescales converge. That is, once the average transition time becomes shorter than the correlation time of the system, the original AMD methods no longer provide any advantage over direct MD.

In order to improve the spatial scaling of traditional AMD methods, new approaches have been introduced based on the concept of localization. Although the global transition rate is likely to grow when a physical system increases in size, the transition rate within a given local region is unlikely to change, because a very large majority of transition processes are local in nature, so that while the absolute number of possible transitions increases with system size, their spatial density remains relatively constant.

In the following section, we discuss two novel methodologies that exploit this locality. First, we discuss the synchronous sub-lattice (SL) algorithm, which has been used to improve the performance of the PRD and TAD approaches. Second, we discuss the recently introduced local hyperdynamics method, which can improve the spatial scaling of hyperdynamics without the need for SL decomposition.

## 5.1    Synchronous Sub-lattice Decomposition

The synchronous sub-lattice (SL) algorithm is a two-layer spatial decomposition routine that was originally designed to improve the spatial scaling of KMC simulations (Shim and Amar 2005). Over the past decade, this approach has also been used to improve the performance of both TAD (Shim et al. 2007) and PRD (Martínez et al. 2014). The recursive subdivision procedure, at the heart of SL, results in a set of noninteracting regions that can be advanced concurrently in time. This ultimately allows multiple computing units to work on distinct localized regions of the sample, in parallel.

In contrast to spatially parallel MD, the approach is designed to reduce the frequency of communication between neighboring regions while retaining a sufficient level of accuracy. Figure 6 shows a two-dimensional SL decomposition schematically, where numbers and letters refer to domains and subdomains, respectively. Starting with a system of atoms, the global geometry is first partitioned into domains, which can each be assigned to distinct processors (numbers in Fig. 6). Each domain is then further divided into a set of $2^{Dim}$ sub-domains (letters in Fig. 6), where $Dim$ is the number of spatial dimensions used in the decomposition. Intra-domain labeling is assigned in a cyclic fashion, so that sub-domains sharing the same letter do not interact. The collection of all sub-domains with a matching label (letter) is called a sub-lattice.

**Fig. 6** Illustration of the synchronous sub-lattice algorithm used in SLPRD and parTAD. Note that the image corresponds to a two-dimensional discretization, requiring four sub-domains for each *processor* domain
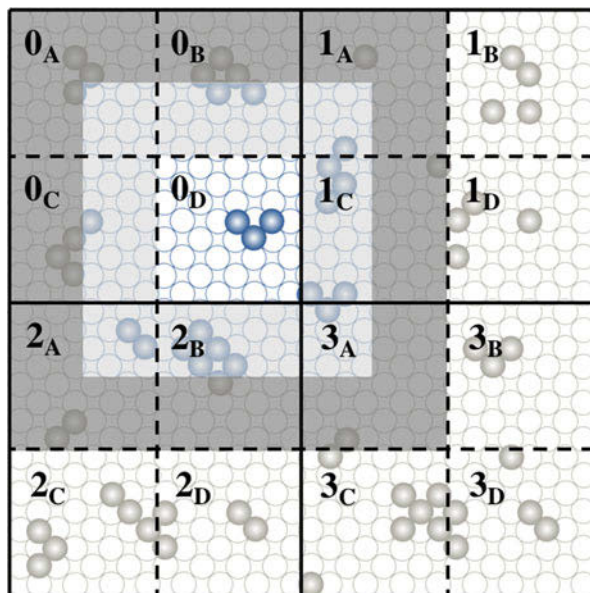
To advance the global simulation time by $\tau_{adv}$ using the SL procedure, each of the $2^{Dim}$ sub-lattices are chosen sequentially, with the full loop resulting in a sub-lattice cycle. For each loop iteration, a desired modeling algorithm (KMC, PRD, TAD, etc.) is used to advance the simulated time by $\tau_{adv}$ within the corresponding sub-lattice. Since the sub-lattice is composed of noninteracting sub-domains, each of these regions can be advanced concurrently by distinct processors.

Since neighboring sub-domains must be synchronized on a timescale that is coarse relative to MD time integration, a boundary skin must be added to each of the sub-domains when the local time is advanced. This skin must include an inner boundary of moving atoms (the light-gray region surrounding sub-domain $0_D$ in Fig. 6), as well as an outer boundary of fixed atoms (the dark-gray region in Fig. 6). Although the inner-boundary atoms are allowed to move when the local time is advanced, the sub-domain must explicitly reject any transitions outside the inner boundary, unless the center of mass (the center of mass of all atoms involved in the transition) is contained within the sub-domain.

The SL algorithm is designed to address boundary conflicts in which two mutually exclusive events occur at some domain boundary. However, it does not avoid causality errors in which events happening at different times might affect the dynamics of the system. Although an extension of the original algorithm exists in which causality errors are completely avoided (Martínez et al. 2008, 2011), some small bias is still introduced by the spatial decomposition itself. In most cases, this bias is known to be relatively small (Shim and Amar 2005; Martínez et al. 2011).

### 5.1.1   Sub-lattice PRD

The sub-lattice PRD method (SLPRD) corresponds to an SL implementation in which PRD is used to advance the simulated time within each sub-domain (Martínez et al. 2014). In general, the parallel speedup of SLPRD is given by

$$\eta^{\text{SLPRD}} = \frac{\tilde{\eta}_{\text{sd}}^{\text{PRD}} D}{\mathcal{O}(skin)},\tag{7}$$

where $D$ is the number of domains , $\tilde{\eta}_{\text{sd}}^{\text{PRD}}$ is the average PRD speedup achieved in each local sub-domain, and $\mathcal{O}(skin)$ is the additional overhead corresponding to the aforementioned boundary regions. The $\mathcal{O}(skin)$ overhead is a linear function of the extra number of atoms in each sub-domain and can be obtained through MD simulations (Martínez et al. 2014). In the hypothetical scenario that $\mathcal{O}(skin) \rightarrow 1$, SLPRD produces an ideal weak scaling of the average PRD boost achieved in each sub-domain. However, the proportion of boundary atoms, with respect to the *active* atoms, is strongly dependent on the size of each processor domain. For this reason, the most efficient use of computing units often requires large SL domains and spatially parallel force calls within each replica process.

In practice, the SLPRD algorithm becomes most advantageous (relative to original PRD with spatially parallel force calls – also known as ParPRD) when there are both many atoms to simulate and many computing units available. For example, Martínez et al. (2014) have directly compared real simulations with a more-detailed form of Eq. (7) to produce the performance projections in Fig. 7. In this figure, MD, ParPRD, and SLPRD are all assumed to leverage spatially parallel force calls. The results clearly show that the SL algorithm becomes most effective when the computing system is extremely large and that the potential for weak scaling is not strongly dependent on the per-atom transition rate of the system.

Although SL does require the computationally expensive use of boundary atoms, the development and analysis of an SLPRD method is certainly a significant AMD advancement. This work has clearly demonstrated both the advantages and limitations of combining SL with a replication-based method like PRD. This accomplishment has ultimately paved the way for the ongoing development of a new sub-lattice ParSplice (SLParSplice) method. Both the theoretical and measured performance of SLPRD, as presented in Martínez et al. (2014) (and reviewed above), can be taken as a lower bound on the available performance of a quality SLParSplice implementation.

### 5.1.2   Sub-lattice TAD

In addition to KMC and PRD, the SL algorithm has also been used within the *parallel* TAD method (parTAD) to dramatically improve the spatial scaling of TAD (Shim et al. 2007). ParTAD is very similar to SLPRD, in the sense that AMD is used to advance the simulation time within each sub-domain of the decomposed sub-lattice geometry. In practice, however, parTAD has an even better chance of achieving improved spatial scaling, because TAD is known to scale especially
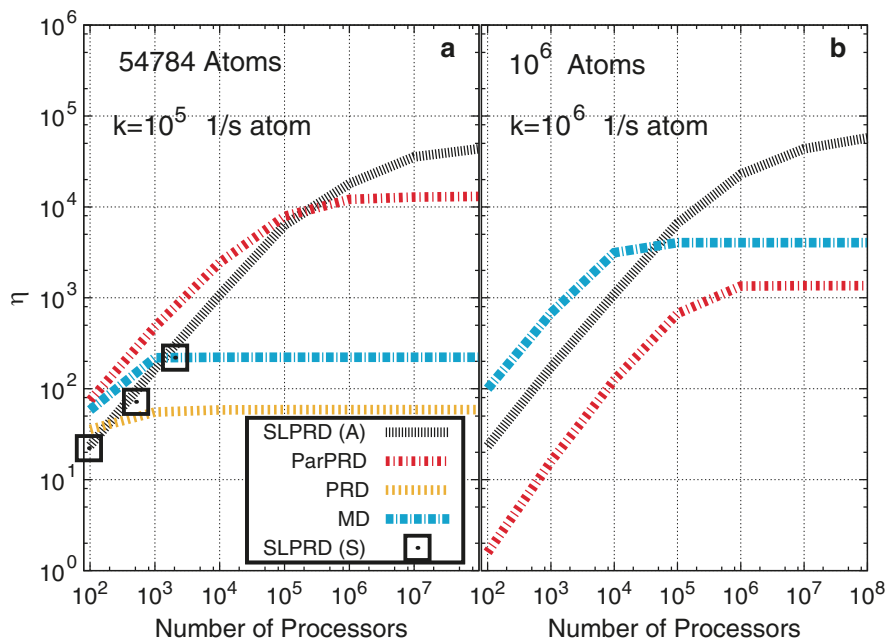
**Fig. 7** Speedup vs. the number of total processors in the system with (**a**) 54,784 atoms and $k = 10^5$ (s atom)$^{-1}$ and (**b**) $10^6$ atoms and $k = 10^6$ (s atom)$^{-1}$. SLPRD(A) and SLPRD(S) show analytical and simulated results, respectively. (Figure reproduced with permission from Martínez et al. 2014)

poorly. That is, for a serial simulation of $N_{at}$ atoms, the run time is expected to scale approximately as $N_{at}^{3+1/3+\gamma}$, where $\gamma = T_{Low}/T_{High}$. This behavior can be improved, to $N_{at}^{2-\gamma}$, when *local* NEB calculations are employed (Shim et al. 2013). However, such a localization must be performed with care, since an undersized NEB geometry can lead to a significant error in the calculated saddle-point energy and therefore the transition rate. Instead of localizing these calculations directly, SL can accomplish the same NEB performance while also reducing the effective transition rate within each domain. Overall, the parallel execution of parTAD can lead to a logarithmic scaling of the serial TAD boost with respect to system size.

## 5.2    Local Hyperdynamics

In hyperdynamics, because the bias potential must go to zero whenever the system is near a dividing surface, increasing the system size will necessarily lead to a lower boost factor for a given definition of the bias potential. Thus, for any proper bias potential, the boost will decay to unity in the limit of large system size. Local hyperdynamics (Kim et al. 2013) mitigates this problem by taking advantage of the

**Fig. 8** Schematic illustration of local hyperdynamics system, showing local domains $I$ and $J$ centered on bonds $i$ and $j$ (shown as dots), respectively. Every bond (or atom) in the system has its own domain. Distant domains do not interact with each other. (Figure reproduced with permission from Kim et al. 2013)
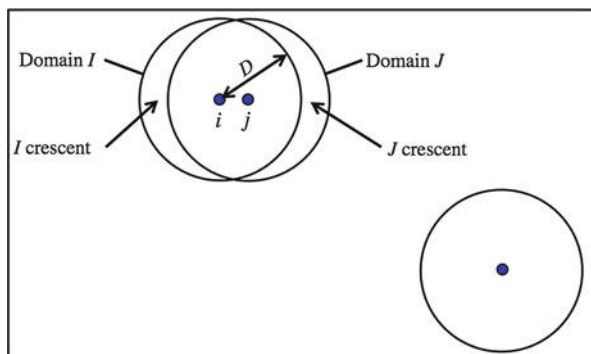


intrinsic locality of the transition pathways. In systems where transitions are local, the dynamics in two regions separated by a large distance evolve independently. More precisely, we define a *locality range L*, beyond which the correlation between bond distortions is zero or smaller than some threshold value set by the user. Given this locality assumption, we define a modified hyperdynamics procedure that only requires that the bias force on an atom be zero when there is a *nearby* geometry distortion taking the system near a dividing surface. We briefly describe the method here, while full detail can be found in Kim et al. (2013).

The local hyperdynamics (LHD) procedure is designed to generate an appropriate escape from the current state, in the same spirit as hyperdynamics, but in a way that maintains a constant boost factor as the system size is increased. In principle, LHD can utilize most any form of bias potential, but for this discussion, we will focus on the use of the simple bond boost (SBB) form, a simplified version (Perez et al. 2009) of the bond boost bias (Miron and Fichthorn 2003), as discussed above. For this type of bias, the natural entity of interest is the bond, rather than the atom. We define a domain $J$, centered on bond $j$, which includes all bonds out to a distance $D$ from $j$, as shown in Fig. 8. Every bond in the system has such a domain associated with it, so that, for example, an $N$-atom fcc system would have roughly $6N$ overlapping domains. For each domain $J$, a scaled version of Eq. (3),

$$V_J^b = V_J^b(\mathbf{R}_J) = C_J V_{\text{SBB}}^b(\mathbf{R}_J), \tag{8}$$

is employed to compute the domain bias energy $V_J^b$, using only the coordinates of the atoms that belong to that domain, as denoted by $R_J$. If, in domain $J$, bond $j$ is the most distorted bond, then the gradient of Eq. (8) is employed to determine the force on bond $j$; otherwise, the force on bond $j$ is zero. (Note that the force on a bond corresponds to the equal, but opposite, force on each atom composing the bond.) Each domain has a bias-energy scaling factor $C_J$; these scaling coefficients are adjusted to give the same average boost factor $B_{\text{target}}$ for every domain in the system. This is accomplished using a *boostostat*, which nudges the coefficient up or down at each time step by an amount proportional to $B_{\text{target}} - B_J$, where $B_J$ is

the instantaneous boost on bond $j$. This is a key component of the LHD procedure, ensuring that all bonds in the system advance in time at the same accelerated rate (on average). We choose a boostostat coupling constant that is not too high (which would distort the dynamics and change the escape rates) and not too low (which would not allow the system to settle into the proper behavior quickly enough relative to the time to next escape. (Note that the best value of $B_{target}$ is system dependent.)

An important characteristic of LHD is that because the force on each bond is governed by a local energy definition, the system dynamics overall is not conservative (i.e., the dynamics with the Langevin thermostat turned off will not be conservative). Surprisingly, though, this procedure nonetheless gives impressively accurate rate constants, typically within a few percent of the exact rates, determined by comparing to MD or regular hyperdynamics benchmarks for various systems. We have shown (Kim et al. 2013) that as the domain size $D$ is increased, the LHD dynamics should become increasingly accurate, even when $D$ is still smaller than the system size. This is because the relative contribution from the crescent regions (see Fig. 8), which give rise to the mismatch between the actual force on bond $j$ and the force on bond $j$ expected by bond $i$, decreases with $D$. (When $D$ is large enough that each domain includes the whole system, then LHD trivially reduces to regular hyperdynamics.) We have also shown (Kim et al. 2013) that for a homogeneous system (all bonds equivalent) and the SBB bias, when $D \geq 2L$, the dynamical evolution will be consistent with Langevin dynamics, in spite of the apparent nonconservative nature, and the accelerated escape rates should be correct. We do not yet know how to show, nor is it necessarily true, that the LHD results are correct for an inhomogeneous system (inequivalent bonds), although the accuracy of the results we have obtained is suggestive.

When a transition from the current state occurs, in addition to redefining the SBB bias potential, adjustment of the $C_J$ coefficients is required for all domains affected by the geometry change caused by the transition. The boostostat, with a properly chosen coupling coefficient, takes care of this, ideally on a timescale that is short compared to the time until the next event in this region. Also, one or more new bonds may have formed due to the transition, requiring creation of new domains and new $C_J$ values, which can be initiated with a good starting guess (e.g., the average $C_J$ value for all bonds in the system), after which the boostostat settles them in.

LHD will give constant boost for an arbitrarily large system, provided that no new lower-barrier pathways show up as the system size is increased; an example for a homogeneous system is shown in Fig. 9. In a massively parallel implementation, LHD will give good parallel weak-scaling provided the extra work required when transitions occur is distributed such that it does not grow with the number of transitions in time.

Although SL could also be used to scale the performance of hyperdynamics, the method requires the use of many boundary atoms to accurately perform MD within each distinct sub-lattice domain. In contrast, the local hyperdynamics approach to localization does not require the use of fixed/free boundary atoms and can therefore be expected to be more efficient than a hypothetical SL-based alternative.
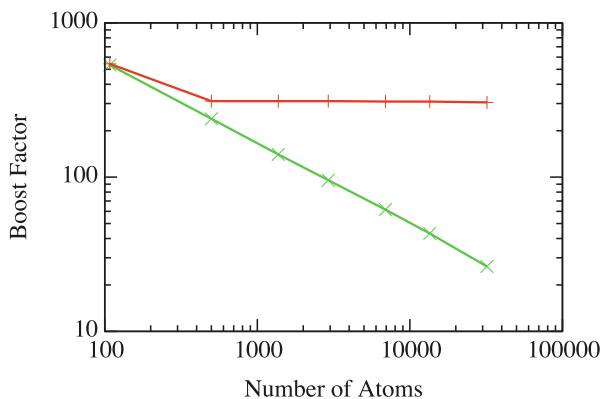
**Fig. 9** Boost factor for local hyperdynamics, red (+), and regular hyperdynamics, green (×) for bulk Ag systems of various sizes modeled with an embedded atom method potential (Voter 1988). (Figure reproduced with permission from Kim et al. 2013)

## 6    Discussion

In this chapter, we reviewed a variety of modern techniques for performing long-timescale AMD simulations using parallel computing systems. The approaches broadly fall into three categories: replication (PRH, ParSplice), speculation (SpecTAD, ParSplice), and localization (SLPRD, parTAD, and local hyperdynamics). Clearly, it is possible for an AMD method to fall into more than one of these categories. For example, the most novel feature of ParSplice is that it can leverage speculative execution to outperform PRD when states are often revisited. However, the ParSplice boost itself relies on a generalization of the replication concept. In fact, the performance of ParSplice is very similar to that of PRD when states are not likely to be repeated, because they are both built on the replication foundation.

Like all the AMD methods that came before them, the primary challenge facing modern AMD algorithms is the so-called *low-barrier* problem. As discussed earlier in the chapter, the boosts of the original PRD, ParSplice, TAD, and hyperdynamics methods degrade dramatically as the global transition rate ($k_g$) increases. When high rates stem from large system sizes, localization methods can be leveraged to lower the effective value of $k_g$.

Unfortunately, many systems have large $k_g$ because the dynamics are simply dominated by low-barrier transition pathways. When these low-barriers connect states into superbasins, modified state definitions or advanced techniques such as SpecTAD and ParSplice can still reach very long timescales. In other words, recent techniques can exploit more general separations of timescales than only between vibrations and transitions. Problems however remain: for example, while it might in principle be possible to define suitably metastable states on complex energy landscapes, doing so in practice remains a considerable challenge that will require new mathematical or physical insights or novel machine-learning

approaches. Further, as the computational scale increases, maintaining efficiency becomes increasingly difficult, as one has to identify a large amount of work that can be carried out concurrently. However, the number of processors that can be efficiently leveraged to generate escapes out of a single or a few states is limited. Efficient speculation then becomes critical, and new developments will be required to make this process efficient.

Although challenges exist, the value of AMD is clearly expanding with the availability of parallel computing resources. In light of new mathematical understandings and rapidly improving parallel algorithms, there is good reason to expect the available time and length scales to continue improving. What is especially exciting, however, is that these AMD advances can (and will) be used to dramatically accelerate the simulation of highly accurate MD-based trajectories. Since computationally expensive force calculations often result in clear opportunities for strong scaling, it has just recently become practical to leverage electronic-structure-based approaches within AMD methodologies. Given that some parallel approaches can be tuned to achieve arbitrary accuracy (with respect to the underlying interatomic forces), it is reasonable to expect that massively parallel AMD will lead to significant materials research progress in the coming years.

# 7   Applications and Lessons Learned

In this chapter, we have focused our attention on the theoretical details of several new parallel AMD methodologies for long-timescale simulations. Since all of the highlighted techniques are still under significant development at the time of writing, some have yet to be used to investigate high-profile materials phenomena. A few of these methods, however, have already proven useful for materials applications of clear technological interest. For example, SpecTAD has been used to investigate the effects of di-vacancy and di-interstitial formation on the mobility of point defects in magnesium aluminate spinel ($MgAl_2O_4$), a complex oxide with applications in both nuclear reactors and military body armor (Zamora et al. 2016b). In $MgAl_2O_4$, the formation of point defect clusters often leads to a complex potential energy landscape in which the system will need to move through a series of deep superbasins for net defect migration to occur. Within each of these superbasins, the energy barriers can be at least an order of magnitude smaller than those needed to escape to external states. For this reason, *synthetic-mode* SpecTAD is extremely effective and can be used to reach years of simulated time at experimentally relevant temperatures. In this case, the long-time simulations provided sufficient simulation data to fully characterize the self-diffusivity of various point defect clusters across a range of possible temperatures.

ParSplice has also proven to be a valuable tool for the investigation of material systems that evolve in time through superbasins of states. Recently, for example, ParSplice was used to simulate the long-time relaxation of platinum nanoparticles, which are of significant interest for possible applications in catalysis and medicine (Perez et al. 2018). By tracking the evolution of an initially disordered

Pt nanoparticle, this work demonstrated that ParSplice can be used to simulate significantly longer timescales than either direct MD or PRD. Since the trajectory is likely to visit many of the same states repetitively, the algorithm can efficiently leverage thousands of cores to produce a simulation rate greater than 1.4 microseconds per minute of WCT (at 300K). For the same physical system, the efficiency of the traditional PRD approach is significantly lower, because the lack of speculation and extensive bookkeeping leads to a much higher dephasing overhead. In fact, this application of ParSplice produced so many state-to-state transitions that a dedicated clustering algorithm was needed to post-process the data (Huang et al. 2017). Using a 16 microsecond (70,000-transition) ParSplice trajectory, the Perron Cluster Cluster Analysis (PCCA) was used to detect a series of ~10 critical superbasin clusters connecting a metastable Marks decahedral state with the thermodynamically favored icosahedral state for a 147-atom nanoparticle.

The possibility of generating large quantities of state-descriptive data is an important similarity between the aforementioned applications of SpecTAD and ParSplice. As a by-product of favorable superbasin simulation performance, these methods can generate state-to-state trajectories that are more difficult to interpret than traditional AMD methods. In order to achieve optimal boost factors, less-important trajectory information (like the details of some intermediate and/or repeated states) is sometimes ignored. For example, the $MgAl_2O_4$ SpecTAD application captured billions of rare events, meaning that information about many of the state visits is necessarily excluded from the raw output. In ParSplice, the generation of splice-able segments is specifically designed to allow shallow intermediate states to be ignored. While it is a clear performance advantage over PRD to allow segments to include multiple intermediate transitions, the properties of a segment can make the final ParSplice trajectory more difficult to parse. Overall, a valuable lesson learned through these recent applications is that high-performance simulation methods tend to require high-performance analysis tools. As parallel algorithms continue to improve, the growing volume and complexity of the raw simulation data is likely to become a key concern for the developers and users of these methods.

# 8 Conclusion

In a world with an increasing availability of distributed computing clusters, the value of parallel algorithm development is continuously growing. As high-performance computing vendors prepare for the exascale and beyond, it is becoming clear that hardware parallelism is likely to increase for the foreseeable future. In this chapter, we have reviewed a variety of modern advances in parallel accelerated molecular dynamics (AMD). To provide a reasonable organization of concepts, we began by introducing the serial AMD approaches and then discussed three distinct classes of parallel methods: replication, speculation, and localization. An emphasis was placed on the more recently developed methods, especially those expected to play a

significant role as the high-performance computing world approaches the exascale. Given the wide range of opportunities and challenges for the application and development of parallel AMD on modern computing resources, there is significant motivation for ongoing and future work.

# References

Audouze C, Massot M, Volz S (2009) Symplectic multi-time step parareal algorithms applied to molecular dynamics. https://hal.archives-ouvertes.fr/hal-00358459

Germann TC, Kadau K (2008) Trillion-atom molecular dynamics becomes a reality. Int J Mod Phys C 19(09):1315–1319

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115(21):9657–9666

Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. J Chem Phys 113(22):9901–9904

Huang R, Lo LT, Wen Y, Voter AF, Perez D (2017) Cluster analysis of accelerated molecular dynamics simulations: a case study of the decahedron to icosahedron transition in pt nanoparticles. J Chem Phys 147(15):152717

Kim SY, Perez D, Voter AF (2013) Local hyperdynamics. J Chem Phys 139(14):144110

Le Bris C, Lelivre T, Luskin M, Perez D (2012) A mathematical formalization of the parallel replica dynamics. Monte Carlo Methods Appl 18(2):119–146

Martínez E, Marian J, Kalos M, Perlado J (2008) Synchronous parallel kinetic Monte Carlo for continuum diffusion-reaction systems. J Comput Phys 227(8):3804–3823

Martínez E, Monasterio P, Marian J (2011) Billion-atom synchronous parallel kinetic Monte Carlo simulations of critical 3D Ising systems. J Comput Phys 230(4):1359–1369

Martínez E, Uberuaga BP, Voter AF (2014) Sublattice parallel replica dynamics. Phys Rev E 89:063308

Miron RA, Fichthorn KA (2003) Accelerated molecular dynamics with the bond-boost method. J Chem Phys 119(12):6210–6216

Niklasson AMN, Mniszewski SM, Negre CFA, Cawkwell MJ, Swart PJ, Mohd-Yusof J, Germann TC, Wall ME, Bock N, Rubensson EH, Djidjev H (2016) Graph-based linear scaling electronic structure theory. J Chem Phys 144(23):234101

Perez D, Uberuaga BP, Shim Y, Amar JG, Voter AF (2009) Accelerated molecular dynamics methods: introduction and recent developments. Ann Rep Comp Chem 5:79–98

Perez D, Uberuaga BP, Voter AF (2015) The parallel replica dynamics method coming of age. Comput Mater Sci 100(Part B):90–103

Perez D, Cubuk ED, Waterland A, Kaxiras E, Voter AF (2016) Long-time dynamics through parallel trajectory splicing. J Chem Theory Comput 12(1):18–28

Perez D, Huang R, Voter AF (2018) Long-time molecular dynamics simulations on massively-parallel platforms: a comparison of parallel replica dynamics and parallel trajectory splicing. J Mater Res 33(7):813822

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comp Phys 117(1):1–19

Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC et al (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 51(7):91–97

Shim Y, Amar JG (2005) Semirigorous synchronous sublattice algorithm for parallel kinetic monte carlo simulations of thin film growth. Phys Rev B 71:125432

Shim Y, Amar JG, Uberuaga BP, Voter AF (2007) Reaching extended length scales and time scales in atomistic simulations via spatially parallel temperature-accelerated dynamics. Phys Rev B 76:205439

Shim Y, Callahan NB, Amar JG (2013) Localized saddle-point search and application to temperature-accelerated dynamics. J Chem Phys 138(9):094101

Smith JS, Isayev O, Roitberg AE (2017) Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. Chem Sci 8:3192–3203

Sørensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112(21):9599–9606

Uberuaga BP, Montalenti F, Germann TC, Voter AF (2005) Accelerated molecular dynamics methods. In: Handbook of materials modeling. Springer Netherlands, Dordrecht, pp 629–648

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3(1):121–127

Voter AF (1988) Simulation of the layer-growth dynamics in silver films: dynamics of adatom and vacancy clusters on ag (100). In: 31st annual technical symposium. International Society for Optics and Photonics, San Diego, pp 214–226

Voter AF (1997) A method for accelerating the molecular dynamics simulation of infrequent events. J Chem Phys 106(11):4665–4677

Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57(22):R13985

Voter AF, Germann TC (1998) Accelerating the dynamics of infrequent events: Combining hyper-dynamics and parallel replica dynamics to treat epitaxial layer growth. In: MRS Proceedings, San Francisco, vol 528

Zamora RJ, Uberuaga BP, Perez D, Voter AF (2016a) The modern temperature-accelerated dynamics approach. Ann Rev Chem Biomol Eng 7(1):3.1–3.24

Zamora RJ, Voter AF, Perez D, Perriot R, Uberuaga BP (2016b) The effects of cation-anion clustering on defect migration in $MgAl_2O_4$. Phys Chem Chem Phys 18:19647–19654

Zamora RJ, Perez D, Voter AF (2018) Speculation and replication in temperature accelerated dynamics. J Mater Res 33(7):823–834

Zepeda-Ruiz LA, Stukowski A, Oppelstrup T, Bulatov VV (2017) Probing the limits of metal plasticity with molecular dynamics simulations. Nature 550(7677):492–495

Tony Lelièvre

## Contents

### Abstract

The objective of this review chapter is to present recent results on the mathematical analysis of the accelerated dynamics algorithms introduced by A.F. Voter in collaboration with D. Perez and M. Sorensen. Using the notion of quasi-stationary distribution, one is able to rigorously justify the fact that the exit event from a metastable state for the Langevin or overdamped Langevin dynamics

T. Lelièvre (✉)
CERMICS, École des Ponts ParisTech, INRIA, Champs-sur-Marne, France
e-mail: tony.lelievre@enpc.fr

can be modeled by a kinetic Monte Carlo model. Moreover, under some geometric assumptions, one can prove that this kinetic Monte Carlo model can be parameterized using Eyring-Kramers formulas. These are the building blocks required to analyze the accelerated dynamics algorithms, to understand their efficiency and their accuracy, and to improve and generalize these techniques beyond their original scope.

# 1    Introduction

The objective of this chapter is to review recent works on the mathematical analysis of the *accelerated dynamics techniques* introduced by A.F. Voter and coworkers from the late 1990s up to recently (Voter 1997, 1998; Sørensen and Voter 2000; Perez et al. 2015a). The objective of these methods is to efficiently simulate thermostated molecular dynamics trajectories over very large timescales. The mathematical analysis which has been developed gives the set of assumptions underlying these algorithms and a way to assess their accuracy. This helps to understand their limitations, but also to improve and extend these numerical methods beyond their original scope.

The aim of statistical computational physics and molecular dynamics is to infer from a microscopic model of matter its macroscopic properties. At the microscopic level, the atomic configuration is given by a set of positions of the atoms (or group of atoms), and the basic ingredient is a potential energy function:

$$V : \mathbb{R}^d \to \mathbb{R}. \tag{1}$$

Here $d$ is the number of degrees of freedom (typically three times the number of atoms), and the function $V$ associates to a given atomic configuration its energy. We will not discuss here how to build such a function $V$. Let us simply mention that this is related to the construction of so-called force fields, which requires a lot of chemical intuition to infer the functional form of the force field and good parametrization using either experimental data or ab initio computations to evaluate the electronic structure associated with given positions of the nuclei. Combining optimally these informations in order to improve the predictive abilities of force fields is a very lively research subject, at the interface between physics, chemistry, numerical analysis, and data sciences.

For a given potential $V$, at a fixed temperature $T$, the configurations of the atomic system at thermal equilibrium are distributed according to the Boltzmann-Gibbs measure (a.k.a. the canonical measure):

$$\mu = Z^{-1} \exp(-\beta V(q)) \, dq \tag{2}$$

where $q \in \mathbb{R}^d$ is the positions of the atoms, $\beta^{-1} = k_B T$ ($k_B$ being the Boltzmann constant), and $Z = \int_{\mathbb{R}^d} \exp(-\beta V(q)) \, dq$ is assumed to be finite. Computing averages with respect to $\mu$ gives access to so-called thermodynamic quantities. Examples include heat capacity, free energy difference, stress, etc. The focus of

this work is rather to discuss numerical methods to compute *dynamical quantities*, namely, observables which depend on the trajectories of the molecular system: transition times, transition paths, reaction mechanisms, etc. This requires to define a dynamics. The typical dynamics one should have in mind is the Langevin dynamics:

$$\begin{cases} dq_t = M^{-1} p_t \, dt \\ dp_t = -\nabla V(q_t) \, dt - \gamma M^{-1} p_t \, dt + \sqrt{2\gamma \beta^{-1}} dW_t \end{cases} \tag{3}$$

where $(q_t, p_t) \in \mathbb{R}^{d \times d}$ denotes the positions and momenta of the particles at time $t \geq 0$, $M$ is the mass tensor, $\gamma > 0$ is a friction parameter, and $W_t$ is a standard $d$-dimensional Brownian motion. Under loose assumptions on $V$, this dynamics is ergodic with respect to the phase-space canonical measure $\overline{Z} \exp(-\beta(V(q) + p^T M^{-1} p/2)) \, dq \, dp$, where $\overline{Z} = \int_{\mathbb{R}^d} \exp(-\beta(V(q) + p^T M^{-1} p/2)) \, dq \, dp$. The marginal in $q$ of this measure is the measure $\mu$ defined above (see (2)). In particular, for any test function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$,

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \varphi(q_s) \, ds = \int_{\mathbb{R}^d} \varphi d\mu. \tag{4}$$

In the following, we will also consider the overdamped Langevin dynamics which is obtained from (3) in the limit $\gamma \to \infty$ or $M \to 0$ (see, e.g., Lelièvre et al. 2010, Section 2.2.4):

$$dX_t = -\nabla V(X_t) \, dt + \sqrt{2\beta^{-1}} dW_t \tag{5}$$

where $X_t \in \mathbb{R}^d$ denotes the positions of the particles. More precisely, the over-damped Langevin dynamics is, for example, derived from the Langevin dynamics in the large friction limit and using a rescaling in time: assuming $M = \text{Id}$ for simplicity, in the limit $\gamma \to \infty$, $(q_{\gamma t})_{t \geq 0}$ converges to $(X_t)_{t \geq 0}$. Again, under loose assumptions on $V$, this dynamics is ergodic with respect to the canonical measure $\mu$. The Langevin and overdamped Langevin dynamics are thus thermostated dynamics: they describe the evolution of the system at a given temperature $T$. Simulating these dynamics over large timescales in order to have access to macroscopic properties of the system (both thermodynamic and dynamical quantities) is the objective of molecular dynamics simulations with applications in many scientific areas: biology, chemistry, materials sciences, etc. To obtain accurate results, this requires to simulate stochastic dynamics in large dimension over very large timescales.

The main difficulty when simulating the dynamics (3) or (5) in practice is that they are metastable. This means that the trajectory of the positions $(q_t)_{t \geq 0}$ or $(X_t)_{t \geq 0}$ remains trapped for very long periods of time in some regions of the configurational space called *metastable states*. This is actually expected from a physical viewpoint: these metastable states typically correspond to some macroscopic conformations of the system. In materials sciences, for example, one could think of these metastable states as positions of some defects in a crystal or of some adatoms on a surface.

In biology, these metastable states can be associated, for example, with molecular conformations of a protein. And it is indeed expected that the residence times in these metastable states are much larger than the typical timescale of vibration within the metastable states. In practice, one needs, for example, to use timesteps of the order of $10^{-15}$ s to discretize (3), while the typical phenomena of interest occur over timescales of the order of $10^{-6}$ s up to seconds or even more! The numerical counterpart is twofold: to compute thermodynamic quantities, the convergence of time averages such as (4) is very slow; and to evaluate dynamical quantities of interest, namely, typically transition paths and transition times between metastable states, one needs to sample rare events, namely, the exit from metastable states and the transition to a new metastable state. The objective of the *accelerated dynamics* algorithms is indeed to efficiently sample metastable dynamics in order to have access to some dynamical quantities. Let us make precise that we are here considering numerical methods to sample the whole dynamics from states to states and not, for example, the ensemble of reactive paths between two given metastable states, for which other dedicated methods can be used (splitting techniques (van Erp et al. 2003; Allen et al. 2005; Cérou et al. 2011), transition path sampling methods (Dellago et al. 1999), etc.)

The basic idea of the *accelerated dynamics* algorithms is the following: if the stochastic process $(q_t, p_t)_{t \geq 0}$ or $(X_t)_{t \geq 0}$ remains trapped for a sufficiently long time in a metastable state, it forgets the way it entered this state, and this means that the exit event from this metastable state can be modeled by the exit event of a kinetic Monte Carlo model, namely, a pure jump Markov process. This will be explained in Sect. 2. We will then show in Sect. 3 that, in the small temperature regime, it is possible to parameterize the kinetic Monte Carlo model modeling the exit event using Eyring-Kramers laws, which gives the basis for a rigorous foundation of the harmonic transition state theory. From these properties, it is then possible to devise efficient algorithms to simulate metastable dynamics. This is explained in Sect. 4. The bottom line is that the details of the dynamics within metastable states are not interesting: only the exit events from these states need to be simulated.

## 2    Kinetic Monte Carlo Models and Quasi-stationary Distribution

In this section, we consider a set $S \subset \mathbb{R}^d$, which is assumed to be bounded, regular, and open. This set is intended to be associated with one state of a kinetic Monte Carlo (kMC) model, and we would like to relate the exit event from $S$ using the Langevin (3) or overdamped Langevin (5) dynamics and a kinetic Monte Carlo model. We will first present in Sect. 2.1 how the exit event from a state is modeled in a kinetic Monte Carlo model. We will then introduce in Sect. 2.2 the notion of quasi-stationary distribution (QSD), which is the basic ingredient to connect the simulation of the exit event from a set $S$ for (3) or (5) with a kMC model, as explained in Sect. 2.3. Finally, we will conclude this section with a discussion on the way to estimate the convergence time to the QSD in Sect. 2.4.

## 2.1 Kinetic Monte Carlo Models

Kinetic Monte Carlo models (a.k.a. Markov state models, see Bowman et al. 2014 and Schütte and Sarich 2013) are continuous-time Markov processes with values in a discrete state space, namely, a jump Markov process (see Voter (2007) for a nice introduction to kMC models). They consist of a collection of states that we can assume to be indexed by integers and rates $(k_{i,j})_{i \neq j \in \mathbb{N}}$ which are associated with transitions between these states. For a state $i \in \mathbb{N}$, the states $j$ such that $k_{i,j} \neq 0$ are the neighboring states of $i$ denoted in the following by:

$$\mathcal{N}_i = \{j \in \mathbb{N}, \, k_{i,j} \neq 0\}. \tag{6}$$

One can thus represent such a jump Markov model as a graph: the states are the vertices, and an oriented edge between two vertices $i$ and $j$ indicates that $k_{i,j} \neq 0$.

Starting at time 0 from a state $Y_0 \in \mathbb{N}$, the model consists in iterating the following two steps over $n \in \mathbb{N}$: Given $Y_n$,

- Sample the residence time $T_n$ in $Y_n$ as an exponential random variable with parameter $\sum_{j \in \mathcal{N}_{Y_n}} k_{Y_n, j}$:

$$\forall t \geq 0, \; \mathbb{P}(T_n \geq t | Y_n = i) = \exp\left(-\left[\sum_{j \in \mathcal{N}_i} k_{i,j}\right] t\right). \tag{7}$$

- Sample independently from $T_n$ the next visited state $Y_{n+1}$ starting from $Y_n$ using the following law:

$$\forall j \in \mathcal{N}_i, \; \mathbb{P}(Y_{n+1} = j | Y_n = i) = \frac{k_{i,j}}{\sum_{j' \in \mathcal{N}_i} k_{i,j'}}. \tag{8}$$

The associated continuous-time process $(Z_t)_{t \geq 0}$ with values in $\mathbb{N}$ defined by:

$$\forall n \geq 0, \; \forall t \in \left[\sum_{m=0}^{n-1} T_m, \sum_{m=0}^{n} T_m\right), \quad Z_t = Y_n \tag{9}$$

(with the convention $\sum_{m=0}^{-1} = 0$) is then a (continuous-time) jump Markov process.

The exit event from the state $i$ is thus modeled by the couple of random variables $(T, Y)$ where:

1. $T$ and $Y$ are independent;
2. $T$ is exponentially distributed with parameter $\sum_j k_{i,j}$;
3. $Y$ takes the value $j \in \mathcal{N}_i$ with probability $\frac{k_{i,j}}{\sum_{j' \in \mathcal{N}_i} k_{i,j'}}$.

The question we would like to address in the following is: when can one use such a model to simulate the exit event from a metastable state for the Markov dynamics (3) or (5)? The cornerstone to make this connection is the notion of quasi-stationary distribution, which will be introduced in Sect. 2.2.

The interest of using a kMC model rather than the Markov dynamics (3) or (5) to model the evolution of the system is twofold. From a modeling viewpoint, new insights can be gained by building such coarse-grained models that are easier to handle. From a numerical viewpoint, the hope is to be able to build the jump Markov process from short simulations of the full-atom dynamics from states to states. Then, once the rates have been defined, it is possible to simulate the system over much larger timescales than the time horizons attained by standard molecular dynamics, either by using directly the jump Markov process or as a support to accelerate molecular dynamics (see, e.g., Voter (1997, 1998) and Sørensen and Voter (2000) and Sect. 4). It is also possible to use dedicated algorithms to extract from the graph associated with the jump Markov process the most important features of the dynamics (e.g., quasi-invariant sets and essential timescales using the large deviation theory of Freidlin and Wentzell 1984), see, e.g., Wales (2003) and Cameron (2014).

## 2.2    Quasi-stationary Distribution

In this section, we focus for simplicity on the overdamped Langevin dynamics (5). Generalizations to the Langevin dynamics (3) are expected to be true and are the subject of works under progress (see, e.g., Nier (2018) for a first result in that direction). Let us consider the first exit time from a fixed set $S \subset \mathbb{R}^d$:

$$T_S = \inf\{t \geq 0, \ X_t \notin S\}.$$

The exit event from $S$ is fully characterized by the couple of random variables:

$$(T_S, X_{T_S}).$$

The basic intuition already mentioned in the introduction is that if the process remains for a sufficiently long time in $S$, it should be possible to model the exit event by the exit event of a kMC model. This naturally leads us to consider the quasi-stationary distribution which is the invariant law for the process conditioned to stay in $S$.

**Definition 1.** A probability measure $\nu_S$ with support in $S$ is called a quasi-stationary distribution (QSD) for the Markov process $(X_t)_{t\geq 0}$ if and only if

$$\forall t > 0, \ \forall A \subset S, \nu_S(A) = \frac{\displaystyle\int_S \mathbb{P}(X_t^x \in A, t < T_S^x) \, \nu_S(dx)}{\displaystyle\int_S \mathbb{P}(t < T_S^x) \, \nu_S(dx)}.$$

In other words, $\nu_S$ is a QSD if when $X_0$ is distributed according to $\nu_S$, for all positive $t$, the law of $X_t$ conditionally to the fact that $(X_s)_{0 \leq s \leq t}$ remains in the state $S$ is still $\nu_S$.

The QSD satisfies three properties which will be crucial in the following. We refer, e.g., to Le Bris et al. (2012) for a proof of these results and to Collet et al. (2013) for more general results on QSDs.

**Proposition 1.** *Let $(X_t)_{t \geq 0}$ follow the dynamics (5) with an initial condition $X_0 \in S$. Then, there exists a probability distribution $\nu_S$ with support in $S$ such that*

$$\lim_{t \to \infty} \mathcal{L}(X_t | T_S > t) = \nu_S, \tag{10}$$

*where for a given $t > 0$, $\mathcal{L}(X_t | T_S > t)$ denotes the law of the random variable $X_t$ conditioned to the event $\{T_S > t\}$. The distribution $\nu_S$ is the QSD associated with S.*

A consequence of this proposition is the existence and uniqueness of the QSD. In some sense, the QSD can thus be seen as the longtime limit of the process conditioned to stay in the state $S$.

Let us now give a second property of the QSD.

**Proposition 2.** *Let $L = -\nabla V \cdot \nabla + \beta^{-1} \Delta$ be the infinitesimal generator of $(X_t)_{t \geq 0}$ (satisfying (5)). Let us consider the first eigenvalue and eigenfunction associated with the adjoint operator $L^* = \mathrm{div}\,(\nabla V + \beta^{-1} \nabla)$ with homogeneous Dirichlet boundary condition:*

$$\begin{cases} L^* u_1 = -\lambda_1 u_1 \text{ on } S, \\ \quad u_1 = 0 \text{ on } \partial S. \end{cases} \tag{11}$$

*The QSD $\nu_S$ associated with S satisfies:*

$$d\nu_S = \frac{u_1(x)\,dx}{\displaystyle\int_S u_1(x)\,dx}$$

*where $dx$ denotes the Lebesgue measure on S.*

The QSD thus has a density with respect to the Lebesgue measure, which is nothing but the ground state of the Fokker-Planck operator $L^*$ associated with the dynamics with absorbing boundary conditions. The existence, uniqueness, and positivity of this ground state are a standard consequence of the ellipticity of the operator $L^*$ and the boundedness of the set $S$.

Finally, the last property of the QSD concerns the exit event from $S$, when $X_0$ is distributed according to $\nu_S$.

**Proposition 3.** *Let us assume that $X_0$ is distributed according to the QSD $\nu_S$ in $S$. Then the law of the couple $(T_S, X_{T_S})$ (namely, the first exit time and the first exit point from $S$) is fully characterized by the following properties:*

1. *$T_S$ is independent of $X_{T_S}$.*
2. *$T_S$ is exponentially distributed with parameter $\lambda_1$ (defined in Eq. (11) above).*
3. *The law of $X_{T_S}$ is given by (Here and in the following, the superscript $\nu_S$ in $\mathbb{E}^{\nu_S}$ or $\mathbb{P}^{\nu_S}$ indicates that the stochastic process starts under the QSD: $X_0 \sim \nu_S$.): for any bounded measurable function $\varphi : \partial S \to \mathbb{R}$:*

$$\mathbb{E}^{\nu_S}(\varphi(X_{T_S})) = -\frac{\displaystyle\int_{\partial S} \varphi\, \partial_n u_1\, d\sigma}{\beta\lambda_1 \displaystyle\int_S u_1(x)\, dx} \tag{12}$$

*where $\sigma$ denotes the Lebesgue measure on $\partial S$ induced by the Lebesgue measure in $\mathbb{R}^d$ and the Euclidean scalar product and $\partial_n u_1 = \nabla u_1 \cdot n$ denotes the outward normal derivative of $u_1$ on $\partial S$.*

Proposition 3 explains the interest of the QSD. Indeed, if the process is initially distributed according to the QSD in $S$ (namely, from Proposition 1, if it remained for a sufficiently long time in $S$), then the exit event from the state $S$ can be modeled using a kinetic Monte Carlo model, since the exit time is exponentially distributed and independent of the exit point. This will be explained in more detail in the next section.

*Remark 1 (From overdamped Langevin to Langevin).* As mentioned above, the existence of the QSD and the convergence of the conditioned process toward the QSD for the Langevin process (3) require extra work compared to the over-damped Langevin process (5). For results in that direction, we refer to the recent manuscript (Nier 2018). The main difficulties are twofold: (i) even if $S$ is a bounded set, the associated ensemble in phase space is not bounded (the velocities are indeed not bounded) and (ii) the Langevin dynamics is not reversible and not elliptic (noise only acts on velocities, not on positions). This implies some difficulties when studying the spectral properties of the infinitesimal generator with absorbing boundary conditions on $\partial S$.

## 2.3    Modeling of the Exit Event from a Metastable State

Using Proposition 1, if the process $(X_t)_{t \geq 0}$ remains sufficiently long in the state $S$, the random variable $X_t$ is approximately distributed according to the QSD $\nu_S$. Then, from Proposition 3, the exit event $(T_S, X_{T_S})$ satisfies the basic properties needed to be modeled by the exit event of a kinetic Monte Carlo model: $T_S$ is independent from $X_{T_S}$ and exponentially distributed.

**Fig. 1** The boundary $\partial S$ of the domain $S$ is divided into four subdomains $(\partial S_i)_{1 \le i \le 4}$, which are the common boundaries with the neighboring states $(S_i)_{1 \le i \le 4}$
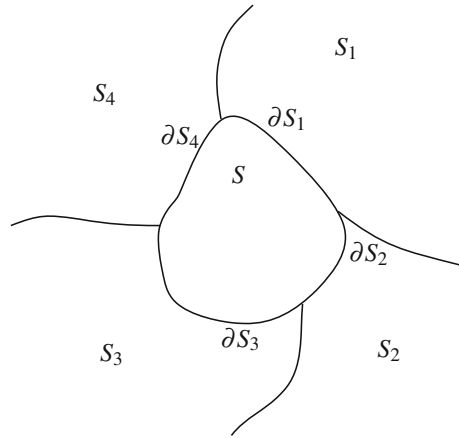


To be more specific, let us consider the domain $S$, and let us divide its boundary $\partial S$ into subsets $(\partial S_i)_{i=1,\dots,J}$ associated with transitions to neighboring states $(S_i)_{i=1,\dots,J}$ (see Fig. 1 for a schematic representation in the case $J = 4$). The next visited state is thus defined by the random variable $Y_S$ with values in $\{1, \dots, J\}$ defined by:

$$Y_S = i \text{ if and only if } X_{T_S} \in \partial S_i.$$

Let us now introduce the rates: for $i = 1, \dots, J$,

$$k_i = \lambda_1 \mathbb{P}^{\nu_S}(X_{T_S} \in \partial S_i) \tag{13}$$

where from (12)

$$\mathbb{P}^{\nu_S}(X_{T_S} \in \partial S_i) = -\frac{\displaystyle\int_{\partial S_i} \partial_n u_1 \, d\sigma}{\beta \lambda_1 \displaystyle\int_S u_1(x) \, dx}. \tag{14}$$

We recall that $(\lambda_1, u_1)$ has been defined in Proposition 2. Now, from Proposition 3, we obviously have the following properties on the couple of random variables $(T_S, Y_S)$:

1. $T_S$ and $Y_S$ are independent.
2. $T_S$ is exponentially distributed with parameter $\sum_{j=1}^{J} k_j$.
3. $Y_S$ takes the value $i \in \{1, \dots, J\}$ with probability $\frac{k_i}{\sum_{j'=1}^{J} k_{j'}}$.

which are exactly the properties needed to model the exit event using a kMC model (see Sect. 2.1).

**Fig. 2** The points $(z_i)_{i=1,\ldots,5}$ are the five local minima of $V$ on $\partial S$. The point $x_1$ is the global minimum of $V$ on $S$. In a global kMC model, exits through the neighborhood of $z_1$ is associated with a transition to $S_1$, exits through the neighborhoods of $z_3$ and $z_4$ are associated with a transition to $S_3$, etc
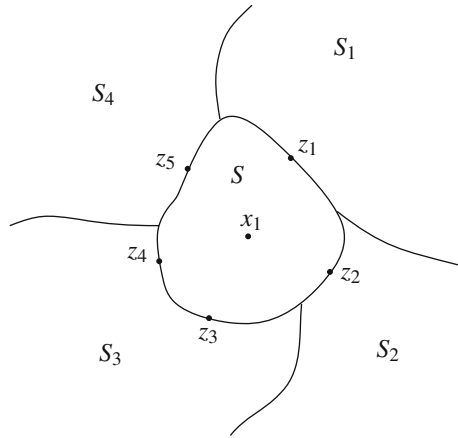
Notice that we have here considered a partition of the boundary $\partial S$ dictated by the a priori knowledge of neighboring states, having in mind a global kMC model as defined in Sect. 2.1. If one is only interested in simulating the exit from $S$ (without any a priori knowledge of $\mathbb{R}^d \setminus S$), a natural partition of the boundary is then

$$\partial S = \cup_{i=1}^I B_{z_i}$$

where $(z_i)_{i=1,\ldots,I}$ are the local minima of $V$ on $\partial S$ (see Fig. 2) and for all $i \in \{1, \ldots, I\}$, $B_{z_i}$ is a neighborhood of $z_i$. Indeed, in the small temperature regime, exits through $\partial S$ occur around the local minima of $V$ on $\partial S$ (this will be discussed in Sect. 3.2, see Remark 3). More precisely, for the mathematical analysis in Sect. 3.2, we will define $B_{z_i}$ as the basin of attraction of $z_i$ for the dynamics $\dot{x} = -\nabla_T V(x)$ in the boundary $\partial S$ (where $\nabla_T V$ denotes the tangential gradient of $V$ along the boundary $\partial S$):

$$B_{z_i} = \left\{ x_0 \in \partial S, \ \lim_{t \to \infty} x(t) = z_i, \ \text{where } x(0) = x_0 \text{ and } \dot{x} = -\nabla_T V(x) \right\}. \quad (15)$$

We will then show that exits through $B_{z_i}$ actually only occur through a neighborhood of $z_i$ in the small temperature regime $\beta \to \infty$. Using this partition, one can thus simulate the exit event as explained above through a couple of random variable $(T_S, Y_S)$, where $Y_S = i$ if the exit occurs through the neighborhood $B_{z_i}$ of $z_i$ (where $i \in \{1, \ldots, I\}$). The associated rates are those defined above, replacing the partition $(\partial S_i)_{i=1,\ldots,J}$ by $(B_{z_i})_{i=1,\ldots,I}$:

$$k_i = \lambda_1 \mathbb{P}^{\nu_S}(X_{T_S} \in B_{z_i}) = -\lambda_1 \frac{\displaystyle\int_{B_{z_i}} \partial_n u_1 \, d\sigma}{\beta \lambda_1 \displaystyle\int_S u_1(x) \, dx}. \quad (16)$$

As explained above, this is useful to simulate the exit from the state $S$ much more efficiently than by integrating in time the Langevin (3) or overdamped Langevin (5) dynamics. It can be used as such in a kinetic Monte Carlo model. It is also the basic ingredient of the accelerated dynamics that will be discussed in Sect. 4. These algorithms aim at accelerating the sampling of the metastable trajectories of the Langevin or overdamped Langevin dynamics, by efficiently generating the exit event from a metastable state $S$ when those dynamics get trapped in $S$.

Notice that the results presented here are very general: they hold for both energetic and entropic traps and for any Markov dynamics as soon as a QSD exists (nonreversible Markov dynamics, discrete-time Markov dynamics, etc.).

One practical difficulty to use the results above as such in a kMC dynamics and in some of the accelerated dynamics algorithms is that one needs in addition a simple way to evaluate the rates $k_i$ (with a more explicit formula than the integral formulation in (16)). We will come back to this in Sect. 3, where the Eyring-Kramers formulas will be introduced to approximate these rates, in the small temperature regime.

## 2.4   Estimating the Convergence Time to the Quasi-Stationary Distribution in Practice

We have seen in the previous section that if the process $(X_t)_{t\geq 0}$ starts from the QSD in $S$, one can model the exit event from $S$ exactly using a kMC model. In addition, in view of Proposition 1, the QSD is reached for the process conditioned to stay in $S$ in the longtime limit. A natural question is then: how long the process $(X_t)_{t\geq 0}$ should remain in $S$ so that one can assume it is sufficiently close to the QSD? In the original paper (Voter 1998), this time is called the correlation time, denoted by $\tau_{\mathrm{corr}}$, and is typically fixed as a constant for all the states, this constant being estimated using some physical intuition on the system at hand or an harmonic approximation. When the dynamics enters a state $S$, one thus waits for some time $\tau_{\mathrm{corr}}$ before assuming that the QSD has been reached: this is called the decorrelation step in Voter (1998). Let us now discuss two ways to estimate the time needed to reach the QSD: one theoretical and one numerical.

From a theoretical viewpoint, the following result proven in Le Bris et al. (2012) gives a first estimate of $\tau_{\mathrm{corr}}$.

**Proposition 4.** *Let* $(X_t)_{t\geq 0}$ *satisfies* (5) *with* $X_0 \in S$. *Let us consider* $-\lambda_2 < -\lambda_1 < 0$ *the first two eigenvalues of the operator* $L^*$ *on* $S$ *with homogeneous Dirichlet boundary conditions on* $\partial S$ *(see Proposition 2 for the definition of* $L^*$*). Then, there exists a constant* $C > 0$ *which depends on the law of* $X_0$*, such that, for all* $t \geq \frac{C}{\lambda_2-\lambda_1}$,

$$\sup_{f,\, \|f\|_{L^\infty}\leq 1} \left| \mathbb{E}(f(T_S - t, X_{T_S})|T_S \geq t) - \mathbb{E}^{\nu_S}(f(T_S, X_{T_S})) \right| \leq C \exp(-(\lambda_2 - \lambda_1)t). \tag{17}$$

In other words, the total variation norm between the law of $(T_S - t, X_{T_S})$ conditioned to $T_S \geq t$ (for any initial condition $X_0 \in S$) and the law of $(T_S, X_{T_S})$ when $X_0$ is distributed according to the QSD $\nu_S$ decreases exponentially fast with rate $\lambda_2 - \lambda_1$. This means that $\tau_{\text{corr}}$ should be chosen of the order $1/(\lambda_2 - \lambda_1)$. There are however two difficulties with this result. First, this is not a very practical result since computing the eigenvalues $\lambda_1$ and $\lambda_2$ is in general impossible. Second, the constant $C$ in the righthand side of (17) and in the inequality $t \geq \frac{C}{\lambda_2 - \lambda_1}$ depends in a complicated way on the law of the initial condition $X_0 \in S$. And one indeed expects the convergence time to the QSD to strongly depend on the initial condition.

A more practical approach to estimate the convergence time $\tau_{\text{corr}}$ to the QSD has been proposed in Binder et al. (2015). The method uses two ingredients:

- The Fleming-Viot particle process (Ferrari and Maric 2007), which consists in $N$ replicas $(X_t^1, \ldots, X_t^N)_{t \geq 0}$ which are evolving and interacting in such a way that the empirical distribution $\frac{1}{N} \sum_{n=1}^N \delta_{X_t^n}$ is close (in the large $N$ limit) to the law of the process $X_t$ conditioned on $t < T_S$.
- The Gelman-Rubin convergence diagnostic (Gelman and Rubin 1992) to estimate the correlation time by the convergence time to a stationary state for the Fleming-Viot particle process.

Roughly speaking, the Fleming-Viot particle process consists in the following: each replica $(X_t^i)_{t \geq 0}$ evolves according to the original dynamics (5) driven by independent Brownian motions (starting from the same initial conditions $X_0$), and, each time one of the replicas leaves the domain $S$, another one taken at random is duplicated. The Gelman-Rubin convergence diagnostic consists in comparing the average of a given observable over replicas at a given time, with the average of this observable over time and replicas: when the two averages are close (up to a tolerance, and for a well-chosen list of observables), the process is considered at stationarity. The interest of this approach is that it gives a practical way to approximate $\tau_{\text{corr}}$ which explicitly takes into account the initial condition of $(X_t)_{t \geq 0}$ in $S$. Of course, one should be cautious when using this technique in practice, since bad choices of the observables may lead to false convergence. We refer to Binder et al. (2015) for examples of applications.

In the following, even though the correlation time $\tau_{\text{corr}}$ depends in principle on the state under consideration $S$ and on the initial condition of the process in $S$, we do not indicate explicitly this dependency and stick to the simple notation $\tau_{\text{corr}}$.

## 3    Eyring-Kramers Law and the Harmonic Transition State Theory

In the previous section, we explained that if the process $(X_t)_{t \geq 0}$ remains sufficiently long in the state $S$, then the exit event can be exactly modeled using a kMC model, with the definitions (16) for the rates associated with exits through local minima of

$V$ on $\partial S$. As will become clear below, some accelerated algorithms will require in addition formulas for the rates, which explicitly depend on the potential $V$ and the temperature $\beta$. This is used in particular in the temperature accelerated dynamics (see Sect. 4.4) to infer the exit event at low temperature from exit events observed at a higher temperature. Such formulas are given by the so-called Eyring-Kramers laws, which are introduced in Sect. 3.1. A mathematical result showing that the exit rates from a state $S$ can indeed be approximated by the Eyring-Kramers laws will then be presented in Sect. 3.2, for the overdamped Langevin dynamics (5).

## 3.1   Eyring-Kramers Laws

The Eyring-Kramers laws give formulas which are used in many contexts as approximations of the rates modeling the exit event from the state $S$ in the small temperature regime ($\beta \to \infty$). Let us consider again the situation of Fig. 2, and let us introduce the global minimum $x_1$ of $V$ in $S$ and the local minima $(z_i)_{i=1,\ldots,I}$ of $V$ on $\partial S$. We assume in the following that they are ordered such that

$$V(z_1) \leq V(z_2) \leq \ldots \leq V(z_I).$$

Notice that we also assume in the following that all the critical points of $V$ and $V|_{\partial S}$ are nondegenerate, which implies in particular that there are a finite number of local minima of $V|_{\partial S}$. The Eyring-Kramers formula gives estimates for the exit rates $(k_j)_{j=1,\ldots,I}$ through neighborhoods of the local minima $(z_j)_{j=1,\ldots,I}$, namely:

$$\forall j \in \{1, \ldots, I\}, \; k_j = \nu_j \exp\big(-\beta[V(z_j) - V(x_1)]\big) \tag{18}$$

where $\nu_j > 0$ is a prefactor which depends on the dynamic under consideration and on $V$ around $x_1$ and $z_j$. Let us give a few examples. If $S$ is taken as the basin of attraction of $x_1$ for the dynamics $\dot{x} = -\nabla V(x)$ so that the points $z_j$ are order one saddle points, the prefactor writes for the Langevin dynamics (3) (assuming $M = \mathrm{Id}$ for simplicity):

$$\nu_j^L = \frac{1}{4\pi} \left( \sqrt{\gamma^2 + 4|\lambda^-(z_j)|} - \gamma \right) \frac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{|\det(\nabla^2 V)(z_j)|}} \tag{19}$$

where $\nabla^2 V$ is the Hessian of $V$ (which, we recall, is assumed to be nondegenerate) and $\lambda^-(z_j)$ denotes the negative eigenvalue of $\nabla^2 V(z_j)$. This formula was derived in Kramers (1940) in a one-dimensional situation. The equivalent formula for the overdamped Langevin dynamics (5) is:

$$\nu_j^{OL} = \frac{1}{2\pi} |\lambda^-(z_j)| \frac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{|\det(\nabla^2 V)(z_j)|}}. \tag{20}$$

Notice that $\lim_{\gamma \to \infty} \gamma \nu_j^L = \nu_j^{OL}$, as expected from the rescaling in time used to go from Langevin to overdamped Langevin (see Sect. 1). The formula (20) has again been obtained in Kramers (1940), but also by many authors previously (see the exhaustive review of the literature reported in Hänggi et al. 1990).

Using the Eyring-Kramers rates (18) in the kinetic Monte Carlo model introduced in Sect. 2.1, one can then model the exit event from $S$ as follows. The exit event from $S$ is modeled by the couple of random variables $(T, Y)$ where $T$ is the first exit time from $S$ and $Y = j$ if the process exits from $S$ in a neighborhood of $z_j$. This couple has the following law:

1. $T$ and $Y$ are independent.
2. $T$ is exponentially distributed with parameter

$$\sum_{j=1}^{n} k_j \sim \left( \sum_{j'=1}^{I_0} \nu_{j'} \right) \exp(-\beta[V(z_1) - V(x_1)]) \tag{21}$$

3. $Y$ takes the value $j \in \{1, \ldots n\}$ with probability

$$\frac{k_j}{\sum_{j'=1}^{n} k_{j'}} \sim \frac{\nu_j}{\sum_{j'=1}^{I_0} \nu_{j'}} \exp(-\beta[V(z_j) - V(z_1)]). \tag{22}$$

where the equivalents $\sim$ are valid in the small temperature regime $\beta \to \infty$ and where $I_0 \in \{1, \ldots, I\}$ denotes the number of global minima of $V$ on $\partial S$:

$$V(z_1) = \ldots = V(z_{I_0}) < V(z_{I_0+1}) \leq \ldots \leq V(z_I).$$

The modeling of the exit event using a kMC model parameterized by the Eyring-Kramers laws is sometimes called the harmonic transition state theory in the literature. We refer, for example, to Voter (2007) for an introduction to this theory and relevant references.

As already explained in Sect. 2.1, using such a model rather than the Langevin or overdamped Langevin dynamics to model the exit event from $S$ is useful either to simulate the evolution of the dynamics over very long time using the kMC model or to accelerate the sampling of metastable trajectories of the Langevin or overdamped Langevin dynamics, as will be explained in Sect. 4.

This raises the following theoretical question: are the Eyring-Kramers rates (18) a valid approximation to model the exit event from $S$ using a kMC model, for the Langevin or overdamped Langevin dynamics? To be more precise, let us consider the overdamped Langevin dynamics (5). We have already seen in Sect. 2.3 that if the stochastic process $(X_t)_{t \geq 0}$ remains for a sufficiently long time in $S$, then it is indeed valid to use a kMC model to simulate the exit event from $S$, with associated rates defined using the eigenvalue-eigenfunction pair $(\lambda_1, u_1)$ (see Eq. (16)). The question is then: can these rates defined by (16) be approximated using the Eyring-Kramers laws (18)?

## 3.2   Approximating the Exit Rates by the Eyring-Kramers Laws: A Mathematical Framework

Let us consider the dynamics (5) with an initial condition distributed according to the QSD $\nu_S$ in a domain $S$. We assume the following:

- The domain $S$ is an open smooth bounded domain in $\mathbb{R}^d$.
- The function $V : \overline{S} \to \mathbb{R}$ is a Morse function with a single critical point $x_1$ (A Morse function is a function such that, at any critical point, the Hessian is nonsingular. Notice that this is in particular required to define the prefactors in the Eyring-Kramers laws.). Moreover, $x_1 \in S$ and $V(x_1) = \min_{\overline{S}} V$.
- The normal derivative $\partial_n V$ is strictly positive on $\partial S$, and $V|_{\partial S}$ is a Morse function with local minima reached at $z_1, \ldots, z_I$ with

$$V(z_1) = \ldots = V(z_{I_0}) < V(z_{I_0+1}) \le \ldots \le V(z_I).$$

- The height of the barrier is large compared to the saddle point height discrepancies: $V(z_1) - V(x_1) > V(z_I) - V(z_1)$.
- For all $i \in \{1, \ldots I\}$, consider $B_{z_i} \subset \partial S$ the basin of attraction for the dynamics in the boundary $\partial S$: $\dot{x} = -\nabla_T V(x)$ (where $\nabla_T V$ denotes the tangential gradient of $V$ along the boundary $\partial S$, see Eq. (15)). We assume that

$$\inf_{z \in B_{z_i}^c} d_a(z_i, z) > \max(V(z_I) - V(z_i), V(z_i) - V(z_1)) \qquad (23)$$

where $B_{z_i}^c = \partial S \setminus B_{z_i}$.

Here, $d_a$ is the Agmon distance:

$$d_a(x, y) = \inf_{\gamma \in \Gamma_{x,y}} \int_0^1 g(\gamma(t))|\gamma'(t)| \, dt$$

where $g = \begin{cases} |\nabla V| \text{ in } S \\ |\nabla_T V| \text{ in } \partial S \end{cases}$, and the infimum is taken over the set $\Gamma_{x,y}$ of all piecewise $C^1$ paths $\gamma : [0, 1] \to \overline{S}$ such that $\gamma(0) = x$ and $\gamma(1) = y$. The Agmon distance is useful in order to measure the decay of eigenfunctions away from critical points. These are the so-called semiclassical Agmon estimates (see Simon 1984 and Helffer and Sjöstrand 1984). Under the assumptions stated above, the following result is proven in Di Gesù et al. (2017).

**Theorem 1.** *Under the assumptions stated above, in the limit* $\beta \to \infty$*, the exit rate is:*

$$\lambda_1 = \sqrt{\frac{\beta \det(\nabla^2 V)(x_1)}{2\pi}} \sum_{k=1}^{I_0} \frac{\partial_n V(z_k)}{\sqrt{\det(\nabla^2 V_{|\partial S})(z_k)}} e^{-\beta(V(z_1)-V(x_1))}(1 + O(\beta^{-1})).$$

$$(24)$$

*Moreover, for any open set $\Sigma_i$ containing $z_i$ such that $\overline{\Sigma}_i \subset B_{z_i}$,*

$$\mathbb{P}^{\nu_S}(X_{T_S} \in \Sigma_i) = \frac{\frac{\partial_n V(z_i)}{\sqrt{\det(\nabla^2 V|_{\partial S})(z_i)}}}{\sum_{k=1}^{I_0} \frac{\partial_n V(z_k)}{\sqrt{\det(\nabla^2 V|_{\partial S})(z_k)}}} e^{-\beta(V(z_i)-V(z_1))}(1 + O(\beta^{-1})). \quad (25)$$

Formula (24) (resp. (25)) should be compared with (21) (resp. (22)) introduced above. We refer to Di Gesù et al. (2017) for a proof and for other related results (see also Helffer and Nier (2006) for a proof of (24)). The proof uses techniques developed in particular in the previous works: Helffer et al. (2004), Helffer and Nier (2006), Le Peutrec (2010), and Lelièvre and Nier (2015). The analysis requires to combine various tools from semiclassical analysis to address new questions: sharp estimates on quasimodes far from the critical points for Witten Laplacians on manifolds with boundary, a precise analysis of the normal derivative on the boundary of the first eigenfunction of Witten Laplacians, and fine properties of the Agmon distance on manifolds with boundary.

Using the two results of Theorem 1 and the formula (16) for the definition of the exit rates, one obtains that the exit rate associated with an exit in the neighborhood $\Sigma_i$ of $z_i$ is

$$k_i = \lambda_1 \mathbb{P}^{\nu_S}(X_{T_S} \in \Sigma_i)$$
$$= \widetilde{\nu}_i^{OL} e^{-\beta(V(z_i)-V(x_1))}(1 + O(\beta^{-1})) \quad (26)$$

where the prefactors $\widetilde{\nu}_i^{OL}$ are given by

$$\widetilde{\nu}_i^{OL} = \sqrt{\frac{\beta}{2\pi}} \partial_n V(z_i) \frac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{\det(\nabla^2 V|_{\partial S})(z_i)}}. \quad (27)$$

This should be compared to the formulas (18) and (20) of the previous section (see Remark 2 below for a discussion on the precise values of the prefactors). This gives a rigorous framework to use a kMC model parameterized using Eyring-Kramers laws to model the exit event from $S$, as introduced in the previous section. Let us emphasize that Theorem 1 provides estimates on the probability $\mathbb{P}^{\nu_S}(X_{T_S} \in \Sigma_i)$ and not only on the logarithm of these probabilities (as obtained, e.g., using large deviation results, see Eq. (28) below). Notice that Theorem 1 also gives error estimates (actually, it can be shown that the terms $O(\beta^{-1})$ in (24), (25), and (26) admit full expansions in positive powers of $\beta^{-1}$).

Let us finish this section with a few important remarks.

*Remark 2 (From generalized saddle points to real saddle points).* As stated in the assumptions, Theorem 1 is proven under the assumption that $\partial_n V > 0$ on $\partial S$: the local minima $z_1, \ldots, z_I$ of $V$ on $\partial S$ are therefore not saddle points of $V$ but so-called generalized saddle points (see Helffer and Nier 2006 and Le Peutrec 2010).

In a future work, we intend to extend these results to the case where the points $(z_i)_{1 \le i \le I}$ are saddle points of $V$ (This is indeed natural if $S$ is taken as the basin of attraction of $x_1$ for the gradient dynamics $\dot{x} = -\nabla V(x)$, as in Sect. 3.1.), in which case we expect to prove the same result (26) for the exit rates, with the prefactor $\widetilde{v}_i^{OL}$ being $\dfrac{1}{\pi} |\lambda^-(z_j)| \dfrac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{|\det(\nabla^2 V)(z_j)|}}$ (this formula can be obtained using formal expansions on the exit time and the Laplace's method). Notice that the latter formula differs from (20) by a multiplicative factor $1/2$. This is due to the fact that $\lambda_1$ is the exit rate from $S$ and not the transition rate to one of the neighboring state (see, e.g., the remark on page 408 in Bovier et al. (2004) on this multiplicative factor $1/2$ and the results on asymptotic exit times in Maier and Stein 1993). One way to understand this multiplicative factor $1/2$ is that once on the saddle point, in the limit $\beta \to \infty$, the process has a probability one half to go back to $S$ and a probability one half to effectively leave $S$. This multiplicative factor does not have any influence on the law of the next visited state which only involves a ratio of the rates $k_i$ (see Eq. (8)).

*Remark 3 (On the importance of prefactors).* The importance of obtaining a result including the prefactors in the rates is illustrated by the following result, which is also proven in Di Gesù et al. (2017). Consider a simple situation with only two local minima $z_1$ and $z_2$ on the boundary $\partial S$, with $V(z_1) < V(z_2)$. Compare the two exit probabilities:

- The probability to leave through $\Sigma_2$ such that $\overline{\Sigma_2} \subset B_{z_2}$ and $z_2 \in \Sigma_2$;
- The probability to leave through $\Sigma$ such that $\overline{\Sigma} \subset B_{z_1}$ and $\inf_\Sigma V = V(z_2)$.

By classical results from the large deviation theory (see, e.g., (28) below), the probability to exit through both $\Sigma$ and $\Sigma_2$ scales like a prefactor times $e^{-\beta(V(z_2)-V(z_1))}$: the difference can only be read from the prefactors. Actually, it can be proven that, in the limit $\beta \to \infty$,

$$\frac{\mathbb{P}^{\nu_S}(X_{T_S} \in \Sigma)}{\mathbb{P}^{\nu_S}(X_{T_S} \in \Sigma_2)} = O(\beta^{-1/2}).$$

The probability to leave through $\Sigma_2$ (namely, through the generalized saddle point $z_2$) is thus much larger than through $\Sigma$, even though the two regions are at the same potential height. This result explains why the local minima of $V$ on the boundary (namely, in our setting the generalized saddle points) play an important role when studying the exit event.

*Remark 4 (On the geometric assumption (23)).* Among the assumptions required to prove Theorem 1, the geometric assumption (23) involving the Agmon distances is probably the most unexpected one. Such an assumption indeed never appeared

before in any mathematical works on Eyring-Kramers laws or on the analysis of the exit event. In Di Gesù et al. (2017), we investigated numerically this assumption, and we observed in some simple geometric settings that if it is not satisfied, the results of Theorem 1 indeed do not hold (more precisely, by computing the exit probabilities (25), we observed that the prefactors are not those predicted by the Eyring-Kramers formulas).

*Remark 5 (From overdamped Langevin to Langevin).* The results of Theorem 1 have been obtained for the overdamped Langevin dynamics (5), and it is then natural to ask whether similar results could be obtained for the Langevin dynamics (3). Indeed such estimates on the rates are actually assumed to be correct for Langevin dynamics in many models and numerical methods using harmonic transition state theory and in particular in the temperature accelerated dynamics algorithm (see Sect. 4.4). We already mentioned in Remark 1 that the existence of the QSD for Langevin requires some additional investigations compared to overdamped Langevin. This is all the more true for the analysis of the exit event starting from the QSD in the small temperature regime. There is hope to obtain similar properties by combining results on the operator associated with Langevin dynamics with absorbing boundary conditions (Nier 2018) and works on the semiclassical analysis of Langevin dynamics (Hérau et al. 2011).

*Remark 6 (On the mathematical results on the Eyring-Kramers law).* Given the importance of the Eyring-Kramers laws in the physics literature, many mathematical approaches have been proposed in order to justify these formulas. Some authors adopt a global approach: they look at the spectrum of the infinitesimal generator of the overdamped Langevin dynamics in the small temperature regime $\beta \to \infty$ (see, e.g., the work by Helffer et al. (2004) based on semiclassical analysis results for Witten Laplacian and the articles by Bovier et al. (2004, 2005) and Eckhoff (2005) where a potential theoretic approach is adopted). These global approaches give the cascade of relevant timescales to reach from a local minimum any other local minimum which is lower in energy. However, they do not give the law of the exit event from a metastable state.

    In the context of this chapter, we are more interested in a local approach: we consider the exit event from a state $S \subset \mathbb{R}^d$, and we would like to relate continuous state-space Markov dynamics such as (5) and (3) and kinetic Monte Carlo model to describe this exit event. In the mathematical literature, the most famous approach to study the exit event is the large deviation theory (Freidlin and Wentzell 1984). In the small temperature regime, large deviation results provide the exponential rates (18), but without the prefactors and without precise error bounds. For the dynamics (5), a typical result on the exit point distribution is the following (see Freidlin and Wentzell 1984, Theorem 5.1): for all $S'$ compactly embedded in $S$, for any $\gamma > 0$, for any $\delta > 0$, there exists $\delta_0 \in (0, \delta]$ and $\beta_0 > 0$ such that for all $\beta > \beta_0$, for all $x \in S'$ such that $V(x) < \min_{\partial S} V$, and for all $y \in \partial S$,

$$\mathrm{e}^{-\beta(V(y)-V(z_1)+\gamma)} \leq \mathbb{P}^x(X_{T_S} \in \mathscr{V}_{\delta_0}(y)) \leq \mathrm{e}^{-\beta(V(y)-V(z_1)-\gamma)} \tag{28}$$

where $\mathscr{V}_{\delta_0}(y)$ is a $\delta_0$-neighborhood of $y$ in $\partial S$. The strength of large deviation theory is that it is very general: it applies to any dynamics (reversible or nonreversible) and in a very general geometric setting, even though it may be difficult in such general cases to make explicit the rate functional and thus to determine the exit probabilities.

Theorem 1 thus differs from all these previous results in the following ways: it gives precise asymptotic estimates of the distribution of $X_{T_S}$; the approach is local, justifies the Eyring-Kramers formula (18) with the prefactors, and provides sharp error estimates.

## 4 Accelerated Dynamics Algorithms

The aim of this section is to explain how the mathematical analysis of metastability and metastable states presented in Sects. 2 and 3 can be used to build efficient algorithms to sample metastable dynamics such as (3) or (5). We present for the sake of simplicity the algorithms in the setting of the overdamped Langevin dynamics (5). The algorithms can be generalized to the Langevin dynamics (3) and actually to many Markov dynamics, under some assumptions that will be discussed below. We will present four algorithms together with their mathematical foundations: parallel replica (Sect. 4.1), parallel trajectory splicing (Sect. 4.2), hyperdynamics (Sect. 4.3), and temperature accelerated dynamics (Sect. 4.4). Finally, we will discuss in Sect. 4.5 how to define in practice the metastable states.

### 4.1 Parallel Replica

The idea of the parallel replica algorithm is to evolve a reference replica following (5) and, if it remains trapped for a long time in a state, to simulate in parallel the exit event from this state. The original parallel replica algorithm (see Voter 1998) consists in iterating three steps:

- *The decorrelation step*: As already explained in Sect. 2.4, in this step, a reference replica evolves according to the original dynamics (5), until it remains trapped for a time $\tau_{\mathrm{corr}}$ in one of the states. During this step, no error is made, since the reference replica evolves following the original dynamics (and there is of course no computational gain compared to a naive direct numerical simulation). Once the reference replica has been trapped in one of the states (that we denote generically by $S$ in the following two steps) for a time $\tau_{\mathrm{corr}}$, the aim is to generate very efficiently the exit event. This is done in two steps.
- *The dephasing step*: In this preparation step, $(N-1)$ configurations are generated within $S$ (in addition to the one obtained from the reference replica) as follows. Starting from the position of the reference replica at the end of the decorrelation step, some trajectories are simulated in parallel for a time $\tau_{\mathrm{corr}}$. For each trajectory, if it remains within $S$ over the time interval of length $\tau_{\mathrm{corr}}$, then its

end point is stored. Otherwise, the trajectory is discarded, and a new attempt to get a trajectory remaining in $S$ for a time $\tau_{\text{corr}}$ is made. This step is pure overhead. The objective is only to get $N$ configurations in $S$ which will be used as initial conditions in the parallel step.

- *The parallel step*: In the parallel step, $N$ replicas are evolved independently and in parallel, starting from the initial conditions generated in the dephasing step, following the original dynamics (5) (with independent driving Brownian motions). This step ends as soon as one of the replica leaves $S$. Then, the simulation clock is updated by setting the residence time in the state $S$ to $N$ (the number of replicas) times the exit time of the first replica which left $S$. This replica now becomes the reference replica, and one goes back to the decorrelation step above.

The computational gain of this algorithm is in the parallel step, which (as explained below) simulates the exit event in a wall clock time $N$ times smaller in average than the wall clock time needed to simulate the exit of the reference walker. This of course requires a parallel architecture able to handle $N$ jobs in parallel. (For a discussion on the parallel efficiency, communication and synchronization, we refer to the papers Voter 1998; Perez et al. 2015b; Le Bris et al. 2012; Binder et al. 2015.) This algorithm can be seen as a way to parallelize in time the simulation of the exit event, which is not trivial because of the sequential nature of time.

In view of the results presented in Sect. 2, the parallel replica is indeed a consistent algorithm (it generates in a statistically correct way the exit event from metastable states) for the following reasons. First, in view of the first property (10) of the QSD, the decorrelation step is simply a way to decide whether or not the reference replica remains sufficiently long in one of the states so that it can be considered as being distributed according to the QSD (see also Sect. 2.4 for an analysis of the error introduced by choosing $\tau_{\text{corr}}$ too small (see in particular Proposition 4) and a discussion on how to evaluate $\tau_{\text{corr}}$ on the fly). Second, by the same arguments, the dephasing step is simply a rejection algorithm to generate many configurations in $S$ independently and identically distributed with law the QSD $\nu_S$ in $S$. Finally, the parallel step generates an exit event which is exactly the one that would have been obtained considering only one replica. Indeed, up to the error quantified in Proposition 4, all the replica are i.i.d. with initial condition, the QSD $\nu_S$. Therefore, according to the third property of the QSD stated in Proposition 3 (see item 1), their exit times $(T_S^n)_{n\in\{1,...N\}}$ are i.i.d. with law an exponential distribution ($T_S^n$ being the exit time of the $n$-th replica) so that

$$N \min_{n\in\{1,...,N\}} (T_S^n) \stackrel{\mathscr{L}}{=} T_S^1. \tag{29}$$

This explains why the exit time of the first replica which leaves $S$ needs to be multiplied by the number of replicas $N$. This also shows why the parallel step gives a computational gain in terms of wall clock: the time required to simulate the exit event is divided by $N$ compared to a direct numerical simulation. Moreover, since

starting from the QSD, the exit time and the exit point are independent (see item 2 in Proposition 3), one also has

$$X^{N_0}_{T_S^{N_0}} \overset{\mathscr{L}}{=} X^1_{T_S^1},$$

where $(X_t^n)_{t \geq 0}$ is the $n$-th replica and $N_0 = \arg\min_{n \in \{1, \ldots, N\}} (T_S^n)$ is the index of the first replica which exits $S$. The exit point of the first replica which exits $S$ is statistically the same as the exit point of the reference walker. Finally, by the independence property of exit time and exit point, one can actually combine the two former results in a single equality in law on couples of random variables, which shows that the parallel step is statistically exact:

$$\left( N \min_{n \in \{1, \ldots, N\}} (T_S^n), X^{N_0}_{T_S^{N_0}} \right) \overset{\mathscr{L}}{=} (T_S^1, X^1_{T_S^1}).$$

We presented the results in the context of the overdamped Langevin dynamics (5) for simplicity. The mathematical analysis actually shows that the parallel replica is a very versatile algorithm. In particular, it can be applied to both energetic and entropic barriers, and it does not assume a small temperature regime (in contrast with the analysis, we will present below for hyperdynamics and temperature accelerated dynamics). In addition, it can also be used for nonequilibrium system (driven by nonconservative forces) and for any Markovian dynamics as soon as a QSD exists. We refer, for example, to Wang et al. (2018) for recent applications to continuous-time Markov chains.

The only errors introduced in the algorithm are related to the rate of convergence to the QSD of the process conditioned to stay in the state. The algorithm is efficient if the convergence time to the QSD is small compared to the exit time (in other words, if the states are metastable). Equation (17) gives a way to quantify the error introduced by the whole algorithm. In the limit $\tau_{\text{corr}} \to \infty$, the algorithm generates exactly the correct exit event. This implies that the resulting dynamics is exact in terms of the laws on trajectories (and not only of the time marginals – master equation – or the stationary state, for example). It gives the correct transition times and trajectories to go from one state to another, for example. The price to pay is that the details of the dynamics within the states are lost.

*Remark 7 (Parallel replica for discrete-time Markov process).* As a remark, let us notice that in practice, discrete-time processes are used (since the Langevin or overdamped Langevin dynamics are discretized in time). Then, the exit times are not exponentially but geometrically distributed. It is however possible to generalize the formula (29) to this setting by using the following fact: if $(\sigma_n)_{n \in \{1, \ldots N\}}$ are i.i.d. with geometric law, then

$$N (\min(\sigma_1, \ldots, \sigma_N) - 1) + \min (n \in \{1, \ldots, N\}, \sigma_n = \min(\sigma_1, \ldots, \sigma_N)) \overset{\mathscr{L}}{=} \sigma_1.$$

We refer to Aristoff et al. (2014) for more details.

*Remark 8 (Generalized parallel replica dynamics).* In view of the analysis above, the crucial parameter of the parallel replica algorithm is the correlation time $\tau_{\text{corr}}$. As explained in Sect. 2.4, Binder et al. (2015) propose an approach to estimate $\tau_{\text{corr}}$ on the fly by combining the Fleming-Viot particle process to simulate the law of the process $X_t$ conditioned on $t < T_S$ and the Gelman-Rubin convergence diagnostic to estimate the convergence time to a stationary state for the Fleming-Viot particle process, which thus gives an estimate of the correlation time. Using this idea, a generalized parallel replica algorithm is introduced in Binder et al. (2015), as follows. Each time the reference replica enters a new state, a Fleming-Viot particle process is launched using $(N - 1)$ replicas simulated in parallel (with initial condition the configuration of the reference replica). Then the decorrelation step consists in the following: if the reference replica leaves $S$ before the Fleming-Viot particle process reaches stationarity, then a new decorrelation step starts (and the replicas generated by the Fleming-Viot particle process are discarded); if otherwise the Fleming-Viot particle process reaches stationarity before the reference replica leaves $S$, then one proceeds to the parallel step. Notice indeed that the final positions of the replicas simulated by the Fleming-Viot particle process can be used as initial conditions for the processes in the parallel step, since they are (approximately) distributed according to the QSD. This procedure thus avoids the choice of a correlation time $\tau_{\text{corr}}$ a priori: it is in some sense estimated on the fly, depending on the state under consideration, and on the initial condition within this state. For more details, discussions on the correlations included by the Fleming-Viot particle process between the replicas and numerical experiments (in particular in cases with purely entropic barriers), we refer to Binder et al. (2015).

## 4.2    Parallel Trajectory Splicing

Parallel trajectory splicing (abbreviated as ParSplice) is a variant of the parallel replica algorithm which has been introduced in Perez et al. (2015a). The idea of the algorithm is to generate in parallel many trajectory segments, which spent at least a time $\tau_{\text{corr}}$ in one state before the beginning of the segment, and ends in a state where it again spends at least a time $\tau_{\text{corr}}$ before the end of the segment. A trajectory segment ending in a state $S$ is then appended to a trajectory segment starting in the same state $S$. The analysis of this algorithm is very similar to the one described above for parallel replica: if $\tau_{\text{corr}}$ is sufficiently large, the trajectory segments start under the QSD in a state and end in a state being again distributed according to the QSD. Therefore, the exit event from the ending state can be simulated by considering any segment which starts under the QSD within the same state. This justifies the fact that these trajectory segments can be appended to form a long molecular dynamics trajectory.

The main difficulty when implementing this technique is to manage the concurrent generation of the segments from many molecular dynamics instances. A database of segments is populated as the simulation proceeds, and this database is used at the same time to generate the molecular dynamics trajectory. This should

be done carefully in order to introduce no bias. For example, segments generated from a given state should not be used in increasing order of their generation time, but using a first-in-first-out queue (otherwise short segments would be favored). In addition, in order to schedule the production of new segments to be added to the database, one uses some prediction of the future trajectory, based on the current knowledge of the visited states. For more details, we refer to Perez et al. (2015a).

## 4.3    Hyperdynamics

As in parallel replica, let us assume that a reference replica $(X_t)_{t \geq 0}$ following the overdamped Langevin dynamics (5) remains trapped for a time $\tau_{\text{corr}}$ is a metastable state $S$. If $\tau_{\text{corr}}$ is sufficiently large, one can assume that the process is distributed according to the QSD. The principle of the hyperdynamics algorithm is then to raise the potential inside the state in order to accelerate the exit from $S$. The algorithm thus requires a biasing potential $\delta V : S \to \mathbb{R}$, which satisfies appropriate assumptions detailed below. The algorithm then proceeds as follows:

- Equilibrate the dynamics on the biased potential $V + \delta V$, namely, run the dynamics (5) on the process $(X_t^{\delta V})_{t \geq 0}$ over the biased potential conditionally to stay in the well, up to the time the random variable $X_t^{\delta V}$ has distribution close to the QSD $\nu_S^{\delta V}$ associated with the biased potential. This first step is a preparation step, which is pure overhead. The end point $X_t^{\delta V}$ will be used as the initial condition for the next step.
- Run the dynamics (5) over the biased potential $V + \delta V$ up to the first exit time $T_S^{\delta V}$ from the state $S$. The simulation clock is updated by adding the effective exit time $B \times T_S^{\delta V}$ where $B$ is the so-called boost factor defined by:

$$B = \frac{1}{T_S^{\delta V}} \int_0^{T_S^{\delta V}} \exp\left(\beta \, \delta V(X_t^{\delta V})\right) dt. \tag{30}$$

The exit point is then used as the starting point for a new decorrelation step.

Roughly speaking, the assumptions required on $\delta V$ in the original paper (Voter 1997) are twofold:

- $\delta V$ is sufficiently small so that the exit event from the state $S$ can be modeled by a kinetic Monte Carlo models parameterized by the Eyring-Kramers laws.
- $\delta V$ is zero on a neighborhood of the boundary $\partial S$.

The derivation of the method relies on explicit formulas for the laws of the exit time and exit point, using the harmonic transition state theory, as explained in Sect. 3.
The algorithm we present here is actually slightly different from the way it is introduced in the original paper (Voter 1997). Indeed, in the original version, the

local equilibration steps (decorrelation step and equilibration step on the biased potential) are omitted: it is assumed that the states are sufficiently metastable (for both the original potential and the biased potential) so that these local equilibrations are immediate. It would be interesting to check if the modifications we propose here improve the accuracy of the method.

Let us now discuss the mathematical foundations of this technique and, in particular, a way to understand the formula (30) for the boost factor. We actually need to compare two exit events. The first one is the exit event for the original process $X_t$ following the dynamics (5), starting from the QSD $\nu_S$ associated with the state $S$ and the dynamics with potential $V$. The second one is the exit event for the process $X_t^{\delta V}$ following the dynamics (5) on the biased potential $V + \delta V$, starting from the QSD $\nu_S^{\delta V}$ associated with the state $S$ and the dynamics with potential $V + \delta V$. Referring to Sect. 3, one way to justify the algorithm is to use the fact that, both on the original potential $V$ and the biased potential $V + \delta V$, a kMC model with transition rates defined by (18) with prefactors such as (19), (20), or (27) can be used to model the exit event (see in particular Theorem 1). Indeed, if this is the case, by using the biasing potential $V + \delta V$, one easily checks that the ratio of exit rates $k_j / k_i$ (which gives the relative probability to exit through a neighborhood of $z_j$ compared to the probability to exit through a neighborhood of $z_i$) does not depend on $\delta V$ since, by assumption, $\delta V$ is zero on $\partial S$: this gives the consistency of the algorithm in terms of the distribution of the first exit point. Concerning the exit time, we know that it is exponentially distributed with parameter $\sum_{i=1}^{I} k_i$, with explicit dependency on $V$ (or $V + \delta V$ for the biased potential) in the small temperature regime thanks again to the formulas (18) (and (19), (20), or (27)). For example, in the framework of Theorem 1, the parameter of the exponential law is $\lambda_1$, where the dependency of $\lambda_1$ on the potential $V$ is explicitly given by formula (24). In particular, one easily checks that (indicating the dependency of $\lambda_1$ on the underlying potential in parenthesis)

$$\frac{\lambda_1(V + \delta V)}{\lambda_1(V)} = \sqrt{\frac{\det\big(\nabla^2(V + \delta V)\big)(x_1)}{\det\big(\nabla^2(V)\big)(x_1)}} e^{\beta \delta V(x_1)}(1 + O(\beta^{-1}))$$

(we again used the fact that $\delta V$ is assumed to be zero on $\partial S$). The righthand side is equal, in the regime $\beta \to \infty$, to $\frac{\int_S \exp(-\beta V)}{\int_S \exp(-\beta(V + \delta V))}(1 + O(\beta^{-1}))$, which is indeed well approximated by the boost factor $B$ (in the limit of a sufficiently large residence time $T_S^{\delta V}$, using the ergodicity of the dynamics). This shows that Theorem 1 applied to $V$ and to $V + \delta V$ justifies the use of hyperdynamics.

Let us mention that in Theorem 1, the relative error on the law of the exit event scales like $O(\beta^{-1})$. It is actually possible to show that the relative error for hyperdynamics scales like $O(\exp(-c\beta))$ for some positive $c$, and to prove this result under less stringent geometric conditions than Theorem 1, see Lelièvre and Nier (2015). Again, generalizing these results to the Langevin dynamics, (3) is an open problem that we are currently investigating (see Remark 5).

Notice that, contrary to the parallel replica method, the hyperdynamics is limited to energetic barriers and a small temperature regime, at least for our mathematical analysis. On the other hand, for very high energetic barriers, hyperdynamics is in principle much more efficient: the parallel replica method only divides the exit time by $N$ (the number of replicas), while for deep wells, the boost factor $B$ is very large.

*Remark 9 (On the biasing potential).* A practical aspect we did not discuss so far is the effective construction of the biasing potential $\delta V$. In the original article (Voter 1997), A.F. Voter proposes a technique based on the Hessian $\nabla^2 V$. Alternatively, a well-known method in the context of materials science is the bond-boost method introduced by Miron and Fichthorn (2003). More recently, some authors proposed to build the biasing potential on the fly, by using adaptive biasing techniques (see, e.g., Tiwary and Parrinello (2013), Bal and Neyts (2015), and Dickson (2017)). Let us also mention the recent variant called *local hyperdynamics* (Kim et al. 2013) which does not enter the previous framework since the bias is actually a nonconservative biasing force (which does not derive from a biasing potential).

## 4.4   Temperature Accelerated Dynamics

Let us finally introduce the temperature accelerated dynamics (TAD) (see Sørensen and Voter 2000). Let us assume again that a reference replica $(X_t)_{t \geq 0}$ following the overdamped Langevin dynamics (5) remains trapped for a time $\tau_{corr}$ in a metastable state $S$, so that one can assume that the process is distributed according to the QSD. The principle of TAD is to increase the temperature (namely, increase $\beta^{-1}$ in (5)) in order to accelerate the exit from $S$. The algorithm consists in:

- Simulating many exit events from $S$ at high temperature, starting from the QSD at high temperature,
- Extrapolating the high-temperature exit events to low-temperature exit events using the Eyring-Kramers law (18).

As for the hyperdynamics algorithm, in the original paper by Sørensen and Voter (2000), no equilibration step is used: it is assumed that the states are sufficiently metastable at both high and low temperatures so that the convergence to the QSD is immediate. In particular, in the simulations at high temperature, the replica is simply bounced back into the state $S$ when it leaves $S$ (and not resampled from the QSD in $S$). Let us now describe more precisely how the extrapolation procedure is made.

Let us consider the exit event from $S$, at a given temperature. Using the results of Sect. 2, the exit event can be modeled using a kMC model with transition rates $(k_i)_{i=1,...,I}$ through the local minima $(z_i)_{i=1,...,I}$ of $V$ on $\partial S$. More precisely, let us consider the process evolving in $S$ which is resampled according to the QSD within $S$ after each exit event. For $i \in \{1, \ldots, I\}$, let us denote by $\tau_i$ the first exit time through a neighborhood of $z_i$ for this process. One readily checks that $\tau_i$ is exponentially distributed with parameter $k_i$ and that the $\tau_i$'s are independent.

Therefore, using the notation of Sect. 2.1, the exit time $T_S$ and exit local minimum $Y_S$ ($Y_S = i$ if the exit point is in $B_{z_i}$) satisfies:

$$(T_S, Y_S) \overset{\mathscr{L}}{=} (\min(\tau_1, \dots, \tau_I), \arg\min(\tau_1, \dots, \tau_I)).$$

Now, using the results of Sect. 3 (see in particular Theorem 1), namely, using the Eyring-Kramers law to parameterize the kMC model, one can compute, in the small temperature regime, the ratios $\frac{k_i^{hi}}{k_i^{lo}}$, where $k_i^{hi}$ (resp. $k_i^{lo}$) denotes the rate at high-temperature $\beta^{hi}$ (resp. low-temperature $\beta^{lo}$). For example, if one considers real saddle points on the boundary, with the prefactors (19) (for Langevin) or (20) (for overdamped Langevin), one obtains:

$$\frac{k_i^{hi}}{k_i^{lo}} \simeq \exp\left(-(\beta^{hi} - \beta^{lo})(V(z_i) - V(x_1))\right). \tag{31}$$

Likewise, in the framework of Theorem 1 (namely, for generalized saddle points), one obtains:

$$\frac{k_i^{hi}}{k_i^{lo}} \simeq \sqrt{\frac{\beta^{hi}}{\beta^{lo}}} \exp\left(-(\beta^{hi} - \beta^{lo})(V(z_i) - V(x_1))\right). \tag{32}$$

Using these formulas, one can infer the exit events at inverse temperature $\beta^{lo}$ from the exit events observed at inverse temperature $\beta^{hi}$, since:

$$(\tau_1^{lo}, \dots, \tau_I^{lo}) \overset{\mathscr{L}}{=} (\Theta^1 \tau_1^{hi}, \dots, \Theta^I \tau_I^{hi}) \tag{33}$$

where

$$\Theta^i = \frac{k_i^{hi}}{k_i^{lo}}$$

is a multiplicative factor constructed from the ratio of the rates. In the equality in law in (33), the random variables $\tau_i^{hi/lo}$ are, as described above, exponential random variables with parameter $k_i^{hi/lo}$. To have analytical formula for the correction factors $\Theta_i$ and make the algorithm practical, the Eyring-Kramers law is assumed to be exact, and one uses in practice $\Theta^i = \exp\left(-(\beta^{hi} - \beta^{lo})(V(x_i) - V(x_0))\right)$, see (31) (or $\Theta^i = \sqrt{\beta^{hi}/\beta^{lo}} \exp\left(-(\beta^{hi} - \beta^{lo})(V(x_i) - V(x_0))\right)$, see (32)).

The TAD algorithm thus consists in running the dynamics at high temperature, observing the exit events through the saddle points on the boundary of the state, and updating the exit time and exit region that would have been observed at low temperature. More precisely, if, at a given time, exits through the saddle points $\{s_1, \dots, s_k\} \subset \{1, \dots, I\}$ have been observed, one computes $\min(\Theta^{s_1} \tau_{s_1}^{hi}, \dots, \Theta^{s_k} \tau_{s_k}^{hi})$ and $\arg\min(\Theta^{s_1} \tau_{s_1}^{hi}, \dots, \Theta^{s_k} \tau_{s_k}^{hi})$ to get the corresponding exit time and exit region at low temperature.

The interest of TAD compared to a brute force saddle point search is that it is not required to observe exits through all the saddle points in order to obtain a statistically correct exit event. Indeed, a stopping criterion is introduced to stop the calculations at high temperature when the extrapolation procedure will not modify anymore the low-temperature exit event (namely, will not modify $\min(\Theta^{s_1}\tau_{s_1}^{hi}, \ldots, \Theta^{s_k}\tau_{s_k}^{hi})$, $\{s_1, \ldots, s_k\} \subset \{1, \ldots, I\}$ being the saddle points discovered up to the time the stopping criterion is fulfilled). This stopping criterion requires to provide some a priori knowledge, typically a lower bound on the barriers $V(z_j) - V(x_1)$ ($j \in \{1, \ldots, I\}$) or a lower bound on the prefactors $\nu_j$ in (18) (see, e.g., Aristoff and Lelièvre (2014) for a discussion). In some sense, TAD can be seen as a clever saddle point search, with a rigorous way to stop the searching procedure.

From the above discussion, the mathematical analysis of the TAD algorithm thus requires to prove that the exit event from $S$ can be modeled using a kMC model parameterized by the Eyring-Kramers formulas. This is exactly the content of Theorem 1 for the overdamped Langevin dynamics (5) with generalized saddle points on $\partial S$. See also Aristoff and Lelièvre (2014), for the case of a one-dimensional potential. The generalization of Theorem 1 to real saddle points on $\partial S$ and to the Langevin dynamics (3) is a work under progress (see Remarks 2 and 5).

Notice that, compared to the hyperdynamics, TAD is really based on the Eyring-Kramers formulas, with relative error terms which scales like $1/\beta$, while for hyperdynamics, one can prove that the error is exponentially small in $\beta$. On the other hand, the interest of TAD compared to the hyperdynamics is that it does not require a biasing potential, which may be complicated to build in some situation.

## 4.5   How to Choose the Metastable States?

Let us finally make a few comments on how the metastable states can be defined in practice. As explained above, the overall efficiency of the accelerated dynamics methods depends on the choice of these metastable states: the algorithms indeed provide an acceleration after assuming that the QSD has been reached in a metastable state. The efficiency of the algorithms thus depends on the choice of the states, which should be metastable regions, so that the stochastic process generically reaches the local equilibrium (the QSD) before leaving the state. How to define the metastable states is a difficult question, very much related to the definition of good reaction coordinates (or good reduced degrees of freedom) for molecular dynamics. Notice that choosing good metastable states also implies being able to estimate the correlation time $\tau_{\text{corr}}$ within each state either from some a priori knowledge or some on the fly estimates, as already explained in Sect. 2.4.

In the original papers by Voter (1997, 1998), Sørensen and Voter (2000), and Perez et al. (2015a), the states are defined as the basins of attraction of the gradient dynamics:

$$\frac{dq}{dt} = -\nabla V(q). \tag{34}$$

For almost every initial conditions, this dynamics converges to a local minimum of $V$: there are thus as many states as local minima of $V$, and the states define (up to a negligible set of points) a partition of the state-space $\mathbb{R}^d$. This means that in practice, a steepest descent is performed at regular time intervals to identify the state in which the system is. One big advantage of this definition is that the states do not need to be defined a priori: they can be numbered as they are discovered by the dynamics, when new local minima of $V$ are identified. Formally, in such a case, one can introduce a map

$$\mathscr{S} \colon \mathbb{R}^d \to \mathbb{N}$$

which to a given position associates a label of a basin of attraction of (34). Accelerated dynamics then aim at efficiently simulating the so-called state-to-state dynamics $(\mathscr{S}(q_t))_{t\geq0}$ or $(\mathscr{S}(X_t))_{t\geq0}$.

For a system with more diffusive or entropic barriers (this is typically the case for biological applications), one could think of defining the states using relevant reaction coordinates (see, e.g., Kum et al. (2004) where the states are defined in terms of the molecular topology of the molecule of interest). This requires to identify the states before starting the simulation.

An important point to make is that since the numerical methods are local in nature, one does not need a partition of the state space to apply these techniques. Two situations can then be considered. One can first define metastable states $(S_i)_{i\geq1}$ which are disjoint open subsets of $\mathbb{R}^d$ (they are called milestones by Faradjian and Elber 2004, target sets, or core sets by Schütte et al. 2011). Each time the process enters one of these states, one checks if the QSD is reached before leaving (this is the decorrelation step, see Sect. 2.4), and then parallel replica, hyperdynamics, or temperature accelerated dynamics can be used to efficiently sample the exit event. Parallel trajectory splicing can also be applied in such a situation. These techniques thus do not require a partition. On the other hand, since one has to simulate the original dynamics outside $\cup_{i\geq1}S_i$, if the time spent outside $\cup_{i\geq1}S_i$ is large, the algorithms are less efficient.

Another possibility is to introduce again an ensemble $(C_i)_{i\geq1}$ of disjoint open subsets of $\mathbb{R}^d$ and to define the state $S_i$ as follows:

$$S_i = \mathbb{R}^d \setminus \cup_{j\neq i} C_j.$$

(The state is then unbounded, which raises some mathematical question about the existence and uniqueness of the QSD, which could be addressed under appropriate assumptions on the growth of the potential $V$ at infinity.) The dynamics then goes as follows: when the process enters one of the core set $C_i$, one considers the exit event from the associated state $S_i$, namely, the next visited core set which is different from $C_i$. In terms of kMC dynamics, this corresponds to projecting the dynamics $(X_t)_{t\geq0}$ (or $(q_t)_{t\geq0}$) onto a discrete state-space dynamics obtained by considering the last visited core set as in Vanden-Eijnden et al. (2008) and Schütte et al. (2011). The interest of such an approach is that the time to reach the QSD starting from

$\partial C_i$ in $S_i$ should be quite small if the $C_i$'s are well chosen. In such a setting, the process thus should decorrelate very quickly before exiting, and an approximation using a global kMC algorithm becomes relevant. However, one drawback is that the description of the underlying continuous state-space dynamics may become poor, especially for very small core sets (since the information of the last visited core set may not be very informative about the actual state of the system).

In summary, one should keep in mind that there is possibly room for improvements in the way the metastable states are defined, compared to the original papers where a partition of the state space in basins of attraction of the gradient dynamics (34) is considered (see also Perez et al. (2015b) for a recent discussion). This is particularly relevant since there are now techniques to approximate the convergence time to the QSD which can be applied for sets which are not basins of attractions of (34) (see Binder et al. (2015) and Sect. 2.4).

## 5    Conclusion and Perspectives

We presented a mathematical analysis of the accelerated dynamics proposed by A.F. Voter and coworkers in order to efficiently generate exit events from metastable states. The analysis is based on the notion of quasi-stationary distribution (QSD), which gives a natural framework to prove that the exit event from a state $S$ for the Langevin or overdamped Langevin dynamics can be modeled by a kinetic Monte Carlo (kMC) model, if the QSD is reached before exiting from $S$. Moreover, under some assumptions, we reported on recent mathematical results which show that the Eyring-Kramers formula can be used to parameterize the kMC model.

From a theoretical viewpoint, we already mentioned above that the mathematical analysis proving that the Eyring-Kramers formulas are good approximations of the exit rates is for the moment restricted to the overdamped Langevin dynamics (5) and to a domain $S$ such that $\partial_n V > 0$ on $\partial S$, which prevents us from considering real saddle points $z_i$ on $\partial S$. The generalization of the results presented in Sects. 2 and 3 to the Langevin dynamics (3) and real saddle points on $\partial S$ is a work in progress. It would also be interesting to investigate Langevin or overdamped Langevin dynamics with nongradient forces (nonequilibrium systems).

From a numerical viewpoint, the mathematical analysis shows the versatility and the limitations of the accelerated dynamics algorithms. The parallel replica and the parallel trajectory splicing algorithms, for example, are very general, and can be applied in many contexts: general Markov dynamics, general definitions of states, etc. Likewise, the principle of the temperature accelerated dynamics algorithm can be used as soon as formulas approximating the exit rates are available (one could think, for example, of versions of TAD where the potential is modified, in the spirit of hyperdynamics but without the assumption that the biasing potential is zero on $\partial S$). Generally speaking, the investigation of the performance of these algorithms to new physical settings and general Markov dynamics is a very promising research direction.

# References

Allen R, Warren P, ten Wolde P (2005) Sampling rare switching events in biochemical networks. Phys Rev Lett 94(1):018104

Aristoff D, Lelièvre T (2014) Mathematical analysis of temperature accelerated dynamics. SIAM Multiscale Model Simul 12(1):290–317

Aristoff D, Lelièvre T, Simpson G (2014) The parallel replica method for simulating long trajectories of markov chains. AMRX 2:332–352

Bal K, Neyts E (2015) Merging metadynamics into hyperdynamics: accelerated molecular simulations reaching time scales from microseconds to seconds. J Chem Theory Comput 11(10):4545–4554

Binder A, Simpson G, Lelièvre T (2015) A generalized parallel replica dynamics. J Comput Phys 284:595–616

Bovier A, Eckhoff M, Gayrard V, Klein M (2004) Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. J Eur Math Soc (JEMS) 6:399–424

Bovier A, Gayrard V, Klein M (2005) Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. J Eur Math Soc (JEMS) 7:69–99

Bowman G, Pande V, Noé F (Eds) (2014) An introduction to Markov state models and their application to long timescale molecular simulation. Springer, Dordrecht

Cameron M (2014) Metastability, spectrum, and eigencurrents of the Lennard-Jones-38 network. J Chem Phys 141(18):184113

Cérou F, Guyader A, Lelièvre T, Pommier D (2011) A multiple replica approach to simulate reactive trajectories. J Chem Phys 134:054108

Collet P, Martínez S, San Martín J (2013) Quasi-Stationary Distributions. Springer, Berlin/Heidelberg

Dellago C, Bolhuis P, Chandler D (1999) On the calculation of reaction rate constants in the transition path ensemble. J Chem Phys 110(14):6617–6625

Di Gesù G, Le Peutrec D, Lelièvre T, Nectoux B (2017) Precise asymptotics of the first exit point density. https://arxiv.org/abs/1706.08728

Dickson B (2017) Overfill protection and hyperdynamics in adaptively biased simulations. J Chem Theory Comput 13(12):5925–5932

Eckhoff M (2005) Precise asymptotics of small eigenvalues of reversible diffusions in the metastable regime. Ann Probab 33(1):244–299

Faradjian A, Elber R (2004) Computing time scales from reaction coordinates by milestoning. J Chem Phys 120(23):10880–10889

Ferrari P, Maric N (2007) Quasi-stationary distributions and Fleming-Viot processes in countable spaces. Electron J Probab 12(24):684–702

Freidlin M, Wentzell A (1984) Random perturbations of dynamical systems. Springer, New York

Gelman A, Rubin D (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):457–472

Hänggi P, Talkner P, Borkovec M (1990) Reaction-rate theory: fifty years after Kramers. Rev Mod Phys 62(2):251–342

Helffer B, Nier F (2006) Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach: the case with boundary. Mémoire de la société mathématique de France 105:1–89

Helffer B, Sjöstrand J (1984) Multiple wells in the semi-classical limit I. Commun Partial Diff Equ 9(4):337–408

Helffer B, Klein M, Nier F (2004) Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach. Mat Contemp 26:41–85

Hérau F, Hitrik M, Sjöstrand J (2011) Tunnel effect and symmetries for Kramers-Fokker-Planck type operators. J Inst Math Jussieu 10(3):567–634

Kim S, Perez D, Voter AF (2013) Local hyperdynamics. J Chem Phys 139(14):144110

Kramers H (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7(4):284–304

Kum O, Dickson B, Stuart S, Uberuaga B, Voter AF (2004) Parallel replica dynamics with a heterogeneous distribution of barriers: application to n-hexadecane pyrolysis. J Chem Phys 121:9808–9819

Le Bris C, Lelièvre T, Luskin M, Perez D (2012) A mathematical formalization of the parallel replica dynamics. Monte Carlo Methods Appl 18(2):119–146

Le Peutrec D (2010) Small eigenvalues of the Neumann realization of the semiclassical Witten Laplacian. Ann Fac Sci Toulouse Math (6) 19(3–4):735–809

Lelièvre T, Nier F (2015) Low temperature asymptotics for quasi-stationary distributions in a bounded domain. Anal PDE 8(3):561–628

Lelièvre T, Rousset M, Stoltz G (2010) Free energy computations: a mathematical perspective. Imperial College Press, UK

Maier R, Stein D (1993) Escape problem for irreversible systems. Phys Rev E 48:931–938

Miron R, Fichthorn K (2003) Accelerated molecular dynamics with the bond-boost method. J Chem Phys 119(12):6210–6216

Nier F (2018) Boundary conditions and subelliptic estimates for geometric Kramers-Fokker-Planck operators on manifolds with boundaries, vol 252. American Mathematical Society, Providence

Perez D, Cubuk E, Waterland A, Kaxiras E, Voter AF (2015a) Long-time dynamics through parallel trajectory splicing. J Chem Theory Comput 12(1):18–28

Perez D, Uberuaga B, Voter AF (2015b) The parallel replica dynamics method – coming of age. Comput Mater Sci 100:90–103

Schütte C, Sarich M (2013) Metastability and Markov state models in molecular dynamics. Courant lecture notes, vol 24. American Mathematical Society, Providence

Schütte C, Noé F, Lu J, Sarich M, Vanden-Eijnden E (2011) Markov state models based on milestoning. J Chem Phys 134(20):204105

Simon B (1984) Semiclassical analysis of low lying eigenvalues, II. Tunneling. Ann Math 120: 89–118

Sørensen M, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112(21):9599–9606

Tiwary P, Parrinello M (2013) From metadynamics to dynamics. Phys Rev Lett 111(23):230602

van Erp T, Moroni D, Bolhuis P (2003) A novel path sampling method for the calculation of rate constants. J Chem Phys 118(17):7762–7774

Vanden-Eijnden E, Venturoli M, Ciccotti G, Elber R (2008) On the assumptions underlying milestoning. J Chem Phys 129(17):174102

Voter AF (1997) A method for accelerating the molecular dynamics simulation of infrequent events. J Chem Phys 106(11):4665–4677

Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57(22): R13985

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, BP Uberuaga (eds) Radiation effects in solids. Springer/NATO Publishing Unit, Netherlands, pp. 1–23

Wales D (2003) Energy landscapes. Cambridge University Press, Cambridge

Wang T, Plechac P, Aristoff D (2018) Stationary averaging for multi-scale continuous time Markov chains using parallel replica dynamics. Multiscale Model. Simul., 16:1–27. https://epubs.siam.org/doi/10.1137/16M1108716

# Temporal Acceleration in Coupled Continuum-Atomistic Methods

# 37

Woo Kyun Kim  and Ellad B. Tadmor

## Contents

### Abstract

In order to speed up molecular simulations, coupled continuum-atomistic methods have been developed in which atomistic resolution is only retained in regions of interest with the rest of the model approximated as a continuum.

W. K. Kim
Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH, USA
e-mail: kimwu@ucmail.uc.edu

E. B. Tadmor (✉)
Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN, USA
e-mail: tadmor@umn.edu

In parallel, there have been efforts to extend the time scale accessible in molecular simulations by filtering out atomic vibrations and focusing on the more interesting dynamics associated with the formation and motion of defects. This article focuses on a current research trend to combine these two complementary approaches into a unified framework that can simultaneously span multiple length and time scales from the microscopic to the macroscopic. As a specific example, the combination of the spatial quasicontinuum (QC) method with the temporal hyperdynamics method to create "hyper-QC" is described.

## 1    Introduction

In typical engineering simulations of macroscopic systems, no limitations are placed on length or time scales. Using high-performance computing, a system being modeled can be as large as needed and can be simulated for as long as needed. For example, continuum finite element (FE) simulations of an entire aircraft during flight, or a crash test of a vehicle, are routinely performed. However, the situation is very different when simulating systems on microscopic length scales. In this case, the atoms constituting the material introduce an inherent length and time scale into the problem; these are the separation between atoms on the order of an Angstrom ($10^{-10}$ m) and atomic vibrations that set a characteristic time on the order of a femtosecond ($10^{-15}$ s). This severely limits the size and times that can be modeled, with most simulations restricted to a tiny fraction of the full system on sub-microsecond time scales.

Two parallel efforts have been pursued to address the length and time scale limitations in microscopic simulations. To overcome the length problem, spatial coarse-graining methods have been introduced in which the atoms in the system are strategically thinned in a manner that retains the physics of interest. This can be traced back to the work of Kirkwood (1935) on the potential of mean force and to earlier work on statistical mechanics. More recently, a class of computational partitioned-domain methods have been introduced in which atomistic resolution is retained in regions of interest (e.g., near a crack tip or dislocation core), while a continuum approximation is used elsewhere (Miller and Tadmor 2009). Thus a coupled continuum-atomistic approach is adopted.

Among the various partitioned-domain methods (many of which are reviewed in Miller and Tadmor 2009; Curtin and Miller 2003; Tadmor and Miller 2011), the quasicontinuum (QC) method of Tadmor and co-workers (Tadmor et al. 1996; Shenoy et al. 1999) is one of the earliest and has been widely adopted in various domains. The original static QC method was extended to study finite temperature effects in dynamics simulations; this approach is referred to as "hot-QC" (Dupuy et al. 2005; Tadmor et al. 2013). An alternative "fully nonlocal" version of the QC method (which we refer to as "cluster-QC" (CQC)) was developed by Knap and Ortiz (2001). In CQC, summation in continuum regions is conducted node-by-node using nodal clusters, as opposed to the original QC method that adopts element-by-element summation using the Cauchy-Born rule (see Sect. 3.1).

In parallel to the spatial coarse-graining efforts, different methods for extending the time scale accessible in microscopic simulations have been pursued. These efforts can be traced back to work on the kinetic Monte Carlo (KMC) method in the 1960s (Voter 2007) and earlier work on rate kinetics and transition state theory (TST). A variety of methods have been introduced since then to accelerate dynamics in microscopic simulations. These include hyperdynamics (Voter 1997), the parallel replica method (Voter 1998), temperature-accelerated dynamics (Sørensen and Voter 2000), on-the-fly KMC (Henkelman and Jónsson 2001), metadynamics (Laio and Parrinello 2002), and diffusive molecular dynamics (Li et al. 2011). Briefly, hyperdynamics adds a bias potential to reduce energy barriers and expedite the escape from potential energy wells. In the parallel replica method, several statistical copies of the original system are simultaneously evolved to increase the probability for transitions. Temperature-accelerated dynamics seeks transition pathways at elevated temperatures. The on-the-fly KMC approach constructs the KMC rate table at each state by searching for escape pathways across saddle points. In metadynamics, the potential energy is modified by adding Gaussian potentials defined by several "collective variables" to discourage revisits to previously sampled configurations. In diffusive molecular dynamics, diffusion is simulated by evolving the occupation probabilities of the mean positions of atomic sites by minimizing a free energy expression based on the Gibbs-Bogoliubov inequality.

The spatial coarse graining and temporal acceleration methods described above have largely been developed separately. A current research focus is to formulate methods that combine these efforts into a single framework to enable simulations of large microscopic systems over long times. In this article, we discuss methods where both length and time scales are extended simultaneously. We focus on the "hyper-QC" method, which is an extension of hot-QC based on hyperdynamics. We also briefly describe a second approach that employs the maximum entropy (max-ent) formalism of Jaynes (1957a, b) within CQC, which we refer to as "finite-temperature CQC." (The term "hot-QC" is sometimes applied to this approach as well, which can be confusing.)

We begin in Sect. 2 with general background on the mechanics of systems of particles including the Hamiltonian formulation and statistical mechanics. The hot-QC approach for spatial coarse graining is discussed in Sect. 3. The integration of temporal acceleration into spatial coarse-graining methods, with emphasis on hyper-QC, is in Sect. 4. This includes a discussion of infrequent events and the hyperdynamics method for accelerating molecular dynamics. We end in Sect. 5 with brief concluding remarks.

## 2  Background: Mechanics of Systems of Particles

### 2.1  Hamiltonian Formulation

Consider a material system consisting of $N$ atoms. The total energy of the system, which is called the "Hamiltonian" and denoted by $\mathcal{H}$, is given by

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \mathcal{V}(\boldsymbol{q}) + \mathcal{K}(\boldsymbol{p}), \tag{1}$$

where $\boldsymbol{q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N)$ and $\boldsymbol{p} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)$ represent the positions and momenta of all atoms, $\mathcal{V}$ is the potential energy due to interactions between atoms (see, e.g., Tadmor and Miller 2011), and $\mathcal{K} = \sum_{i=1}^{N} ||\boldsymbol{p}_i||^2/(2m_i)$ is the kinetic energy, where $m_i$ is the mass of particle $i$. It is assumed that the potential energy can be represented as a sum over individual atom energies $V_i$:

$$\mathcal{V} = \sum_{i=1}^{N} V_i(\boldsymbol{q}). \tag{2}$$

Then, the Hamiltonian can also be written as a sum over individual atom Hamiltonians $h_i$:

$$\mathcal{H} = \sum_{i=1}^{N} h_i(\boldsymbol{q}, \boldsymbol{p}_i), \tag{3}$$

where $h_i = V_i(\boldsymbol{q}) + ||\boldsymbol{p}_i||^2/(2m_i)$. Note that the potential energy of atom $i$ will in general depend on some subset of positions in $\boldsymbol{q}$ in the vicinity of $\boldsymbol{q}_i$, whereas the kinetic energy of atom $i$ only depends on $\boldsymbol{p}_i$.

The dynamical trajectories of the atoms are governed by Newton's equations of motion, which in terms of the Hamiltonian are given by

$$\dot{\boldsymbol{q}}_i = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}_i}, \qquad \dot{\boldsymbol{p}}_i = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}_i}, \tag{4}$$

where the dot represents differentiation with respect to time. The first equation in Eq. (4) enforces the connection between momentum and position, and the second is Newton's equation of motion (second law). The dynamical trajectories of all atoms $(\boldsymbol{q}(t), \boldsymbol{p}(t))$ are obtained by integrating the equations of motion subject to any existing macroscopic constraints. This is the method of "molecular dynamics" (MD).

## 2.2   Statistical Mechanics

In practice, it is not possible to integrate the equations of motion in Eq. (4) for a macroscopic system due to the huge numbers of atoms involved, due to lack of knowledge of initial conditions, and due to the exponential sensitivity of trajectories to these conditions. Instead in statistical mechanics, a probabilistic approach is adopted. The probability of finding the system in microstate $(\boldsymbol{q}, \boldsymbol{p})$ is given by a distribution function $\rho(\boldsymbol{q}, \boldsymbol{p})$ satisfying macroscopic constraints. Any macroscopic observable $A$ is then represented as an ensemble average of a corresponding phase function $\mathcal{A}(\boldsymbol{q}, \boldsymbol{p})$:

$$A = \langle \mathcal{A}(\boldsymbol{q}, \boldsymbol{p}) \rangle = \iint \mathcal{A}(\boldsymbol{q}, \boldsymbol{p}) \, \rho(\boldsymbol{q}, \boldsymbol{p}) \, d\boldsymbol{q} \, d\boldsymbol{p}, \tag{5}$$

where the integration is over the entire phase space of the system (i.e., all possible values of $\boldsymbol{q}$ and $\boldsymbol{p}$) and $d\boldsymbol{q} \, d\boldsymbol{p} = d\boldsymbol{q}_1 \ldots d\boldsymbol{q}_N d\boldsymbol{p}_1 \ldots d\boldsymbol{p}_N$. The distribution function $\rho$ satisfies the normalization condition:

$$\iint \rho(\boldsymbol{q}, \boldsymbol{p}) \, d\boldsymbol{q} \, d\boldsymbol{p} = 1. \tag{6}$$

Under equilibrium conditions, for a system of $N$ atoms at constant volume $V$ and in contact with a heat bath at temperature $T$ (the so-called canonical or $NVT$ ensemble), the distribution function has the Boltzmann form:

$$\rho(\boldsymbol{q}, \boldsymbol{p}; \beta) = \frac{1}{\mathcal{Z}(\beta)} e^{-\beta \mathcal{H}(\boldsymbol{q}, \boldsymbol{p})}, \tag{7}$$

where $\beta = (k_B T)^{-1}$, $k_B$ is the Boltzmann constant, and the canonical partition function is

$$\mathcal{Z}(\beta) = \iint e^{-\beta \mathcal{H}(\boldsymbol{q}, \boldsymbol{p})} \, d\boldsymbol{q} \, d\boldsymbol{p}. \tag{8}$$

## 2.3  Harmonic Approximation

The partition function in Eq. (8) can be evaluated approximately by employing the quasi-harmonic (QH) approximation, whereby the potential energy $\mathcal{V}(\boldsymbol{q})$ (which is part of $\mathcal{H}(\boldsymbol{q}, \boldsymbol{p})$) is expanded to second order about mean atom positions $\bar{\boldsymbol{q}}$:

$$\mathcal{V}_{\mathrm{QH}}(\boldsymbol{q}; \bar{\boldsymbol{q}}) = \mathcal{V}(\bar{\boldsymbol{q}}) + \frac{1}{2}(\boldsymbol{q} - \bar{\boldsymbol{q}})^{\mathrm{T}} \mathbf{H}(\boldsymbol{q})(\boldsymbol{q} - \bar{\boldsymbol{q}}). \tag{9}$$

Here $\mathbf{H}$ is the $3N \times 3N$ Hessian (stiffness) matrix, and the positions $\bar{\boldsymbol{q}}$ are free parameters to be determined later. At sufficiently low temperatures (below half the melting temperature according to LeSar et al. 1989), the off-diagonal terms in $\mathbf{H}$ can be neglected, which leads to the local harmonic (LH) approximation:

$$\mathcal{V}_{\mathrm{LH}}(\boldsymbol{q}; \bar{\boldsymbol{q}}) = \mathcal{V}(\bar{\boldsymbol{q}}) + \sum_{i=1}^{N} \frac{1}{2}(\boldsymbol{q}_i - \bar{\boldsymbol{q}}_i)^{\mathrm{T}} \boldsymbol{\Phi}_i(\boldsymbol{q})(\boldsymbol{q}_i - \bar{\boldsymbol{q}}_i), \tag{10}$$

where $\boldsymbol{\Phi}_i$ is the force constant matrix for atom $i$ given by

$$\boldsymbol{\Phi}_i = \begin{bmatrix} \dfrac{\partial^2 \mathcal{V}}{\partial q_{i1} \partial q_{i1}} & \dfrac{\partial^2 \mathcal{V}}{\partial q_{i1} \partial q_{i2}} & \dfrac{\partial^2 \mathcal{V}}{\partial q_{i1} \partial q_{i3}} \\[2mm] \dfrac{\partial^2 \mathcal{V}}{\partial q_{i2} \partial q_{i1}} & \dfrac{\partial^2 \mathcal{V}}{\partial q_{i2} \partial q_{i2}} & \dfrac{\partial^2 \mathcal{V}}{\partial q_{i2} \partial q_{i3}} \\[2mm] \dfrac{\partial^2 \mathcal{V}}{\partial q_{i3} \partial q_{i1}} & \dfrac{\partial^2 \mathcal{V}}{\partial q_{i3} \partial q_{i2}} & \dfrac{\partial^2 \mathcal{V}}{\partial q_{i3} \partial q_{i3}} \end{bmatrix}, \tag{11}$$

and $\boldsymbol{q}_i = (q_{i1}, q_{i2}, q_{i3})$. The LH approximations for the canonical distribution function and partition function follow as

$$\rho_{\mathrm{LH}}(\boldsymbol{q}, \boldsymbol{p}; \beta, \bar{\boldsymbol{q}}) = \frac{1}{\mathcal{Z}_{\mathrm{LH}}} e^{-\beta \mathcal{V}(\bar{\boldsymbol{q}})} \prod_{i=1}^{N} e^{-\beta \left\{ \frac{(\boldsymbol{q}_i - \bar{\boldsymbol{q}}_i)^T \boldsymbol{\Phi}_i (\boldsymbol{q}_i - \bar{\boldsymbol{q}}_i)}{2} + \frac{\|\boldsymbol{p}_i - \bar{\boldsymbol{p}}_i\|^2}{2m_i} \right\}}, \tag{12}$$

$$\mathcal{Z}_{\mathrm{LH}}(\beta, \bar{\boldsymbol{q}}) = e^{-\beta \mathcal{V}(\bar{\boldsymbol{q}})} \prod_{i=1}^{N} \sqrt{\frac{1}{\det \boldsymbol{\Phi}_i}} \left( \sqrt{\frac{2\pi}{\beta}} \right)^3 \left( \sqrt{\frac{2\pi m_i}{\beta}} \right)^3. \tag{13}$$

The identification of $\bar{\boldsymbol{q}}$ with the mean positions of the atoms is confirmed since

$$\langle \boldsymbol{q}_i \rangle_{\mathrm{LH}} = \bar{\boldsymbol{q}}_i \quad (i = 1, \ldots, N). \tag{14}$$

The internal energy and Helmholtz free energy can also be calculated:

$$U_{\mathrm{LH}}(\beta, \bar{\boldsymbol{q}}) = -\frac{\partial \ln \mathcal{Z}_{\mathrm{LH}}}{\partial \beta} = \mathcal{V}(\bar{\boldsymbol{q}}) + \frac{3N}{\beta} = \sum_{i=1}^{N} \left( V_i(\bar{\boldsymbol{q}}) + \frac{3}{\beta} \right), \tag{15}$$

$$\Psi_{\mathrm{LH}}(\beta, \bar{\boldsymbol{q}}) = -\frac{1}{\beta} \ln \mathcal{Z}_{\mathrm{LH}} = \sum_{i=1}^{N} \psi_i(\beta, \bar{\boldsymbol{q}}), \tag{16}$$

where

$$\psi_i(\beta, \bar{\boldsymbol{q}}) = V_i(\bar{\boldsymbol{q}}) + \frac{1}{\beta} \ln \frac{\sqrt{\det \boldsymbol{\Phi}_i(\bar{\boldsymbol{q}})}}{(2\pi \sqrt{m_i}/\beta)^3}. \tag{17}$$

For a quasistatic process, $\bar{\boldsymbol{q}}$ is determined by minimizing $\Psi_{\mathrm{LH}}$, i.e., by solving

$$\frac{\partial \Psi_{\mathrm{LH}}}{\partial \bar{\boldsymbol{q}}_i} = \frac{\partial}{\partial \bar{\boldsymbol{q}}_i} \left( \sum_{i=1}^{N} \psi_i \right) = 0, \quad (i = 1, \ldots, N), \tag{18}$$

and ensuring the solution corresponds to a minimum.

## 3      Spatial Coarse Graining

The atomistic methods described in the previous section can in principle be used to obtain the dynamics of a system of atoms (Sect. 2.1) and its equilibrium properties (Sects. 2.2 and 2.3). However in practice, such simulations become prohibitively expensive when the number of atoms is too large. Spatial multiscale methods, like the hot-QC method (Dupuy et al. 2005; Tadmor et al. 2013), address this by systematically reducing the number of atoms in a manner that retains the physics of interest – a process referred to as "coarse graining." A description of the hot-QC approach to coarse graining follows.

### 3.1      Finite-Temperature Quasicontinuum (Hot-QC)

As a first step to developing a rigorous coarse-graining procedure, one must introduce a well-defined design principle. In the static QC method, the objective is to approximate the total potential energy of the fully atomistic system in a manner that ensures convergence to the exact result as the atomistic region is increased in size (Tadmor et al. 1996). For a dynamical method at finite temperature, a different design principle is required. Hot-QC assumes equilibrium thermodynamic conditions and sets as its goal to approximate phase averages of the system in a controlled fashion. The exact expression being approximated is provided below after some preliminary definitions.

Consider an atomistic system of $N$ atoms that is partitioned into an atomistic region where the deformation changes appreciably on atomic length scales (this is where the interesting physics is occurring) and a continuum region where the deformation varies slowly on the atomic scale (Fig. 1a). The atoms in each region are distinguished using the superscripts "at" and "c" so that $q = (q^{at}, q^c) = (q^{at}_1, \ldots, q^{at}_{N^{at}}, q^c_1, \ldots, q^c_{N^c})$, where $N^{at}$ and $N^c$ are the number of atoms in the atomistic and continuum regions ($N = N^{at} + N^c$). Note that at this stage no coarse graining has taken place.

Returning to the hot-QC design principle, equilibrium properties of this system in the canonical ensemble are given by (Sect. 2.2):

$$A = \frac{1}{\mathcal{Z}} \iiiint \mathcal{A}(q^{at}, p^{at}) e^{-\beta \mathcal{H}(q^{at}, q^c, p^{at}, p^c)} \, dq^{at} \, dp^{at} \, dq^c \, dp^c, \qquad (19)$$

where the dependence on the atoms in the atomistic and continuum regions has been made explicit. Since the intent is to coarse grain the system by removing the atoms in the continuum region, only phase functions that depend on the atoms retained in the atomistic region are considered in Eq. (19). It is straightforward to show that, by reordering the sequence of integrations, Eq. (19) is *exactly* given by the following expression (Shenoy et al. 1999; Tadmor et al. 2013):
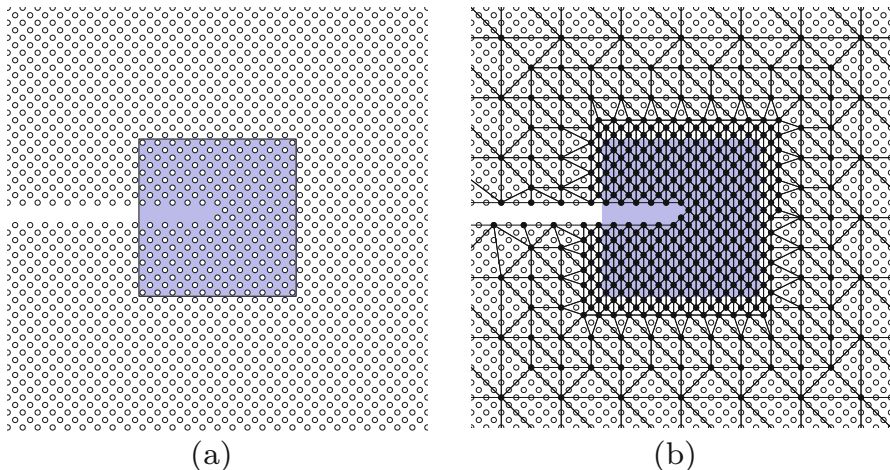
**Fig. 1** A portion of a QC model for an atomically sharp crack. Frame (**a**) shows the atoms in the vicinity of the crack tip. The colored square is the "atomistic region" of QC. The continuum region encompasses all atoms outside this square and extends beyond the boundaries of the figure. Frame (**b**) shows the corresponding QC model. Atoms retained in the model are called "repatoms" and shown as filled circles. The repatoms serve as the nodes of an FE mesh. The positions of unfilled atoms are then obtained through FE interpolation

$$A = \frac{1}{\mathcal{Z}} \iint \mathcal{A}(\boldsymbol{q}^{\text{at}}, \boldsymbol{p}^{\text{at}}) e^{-\beta \mathcal{H}^{\text{at}}(\boldsymbol{q}^{\text{at}}, \boldsymbol{p}^{\text{at}}; \beta)} \, d\boldsymbol{q}^{\text{at}} \, d\boldsymbol{p}^{\text{at}}, \tag{20}$$

where $\mathcal{H}^{\text{at}}$ is the effective Hamiltonian of the atomistic region

$$\mathcal{H}^{\text{at}}(\boldsymbol{q}^{\text{at}}, \boldsymbol{p}^{\text{at}}; \beta) = \mathcal{V}^{\text{at}}(\boldsymbol{q}^{\text{at}}; \beta) + \mathcal{K}^{\text{at}}(\boldsymbol{p}^{\text{at}}; \beta), \tag{21}$$

and $\mathcal{V}^{\text{at}}$ and $\mathcal{K}^{\text{at}}$ are the corresponding effective potential and kinetic energies:

$$\mathcal{V}^{\text{at}}(\boldsymbol{q}^{\text{at}}; \beta) = -\frac{1}{\beta} \ln \int e^{-\beta \mathcal{V}(\boldsymbol{q}^{\text{at}}, \boldsymbol{q}^{\text{c}})} \, d\boldsymbol{q}^{\text{c}}, \tag{22}$$

$$\mathcal{K}^{\text{at}}(\boldsymbol{p}^{\text{at}}; \beta) = -\frac{1}{\beta} \ln \int e^{-\beta \mathcal{K}(\boldsymbol{p}^{\text{at}}, \boldsymbol{p}^{\text{c}})} \, d\boldsymbol{p}^{\text{c}}. \tag{23}$$

The effective potential energy $\mathcal{V}^{\text{at}}$ corresponds to the potential of mean force introduced by Kirkwood (1935).

The hot-QC approach is based on an approximate computation of $\mathcal{H}^{\text{at}}$. The effective kinetic energy in Eq. (23) can be computed analytically. (See Dupuy et al. (2005) for details on how coarse graining is handled for this term.) The effective potential energy $\mathcal{V}^{\text{at}}$ in Eq. (22) can be approximated by expanding the potential energy about mean positions of the continuum atoms $\bar{\boldsymbol{q}}^{\text{c}}$ and applying the LH approximation (Sect. 2.3):

$$\mathcal{V}_{\text{LH}}^{\text{at}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}; \beta) = \sum_{i=1}^{N^{\text{at}}} V_i^{\text{at}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}) + \sum_{i=1}^{N^{\text{c}}} \psi_i^{\text{c}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}; \beta), \qquad (24)$$

where $V_i^{\text{at}}$ is the potential energy of atom $i$ in the atomistic region and $\psi_i^{\text{c}}$ follows in similar fashion to Eq. (17):

$$\psi_i^{\text{c}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}; \beta) = V_i^{\text{c}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}) + \frac{1}{2\beta} \ln \left[ \det \boldsymbol{\Phi}_i^{\text{c}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}) \right], \qquad (25)$$

where $V_i^{\text{c}}$ refers to the potential energy of continuum atom $i$. (Note that $\psi_i^{\text{c}}$ only includes the potential part of the free energy, since the kinetic part is included in $\mathcal{K}^{\text{at}}$.) The LH expression in Eq. (24) depends on the mean positions of the continuum atoms. Treating the continuum as a heat bath that occupies an equilibrium state as the atomistic region evolves, $\mathcal{V}^{\text{at}}$ in Eq. (22) is identified with the minimum of the LH expression:

$$\mathcal{V}^{\text{at}}(\boldsymbol{q}^{\text{at}}; \beta) \approx \min_{\bar{\boldsymbol{q}}^{\text{c}}} \mathcal{V}_{\text{LH}}^{\text{at}}(\boldsymbol{q}^{\text{at}}, \bar{\boldsymbol{q}}^{\text{c}}; \beta). \qquad (26)$$

The expression in Eq. (24) still involves a sum over all $N^{\text{c}}$ atoms in the continuum region making it prohibitively expensive. To reduce the computational cost, a small subset of atoms is selected from the continuum region to represent the remainder. The set of all atoms in the atomistic region and the selected atoms in the continuum region are called "representative atoms" or "repatoms" for short. The atoms in the atomistic region are "nonlocal" repatoms (since their energy depends nonlocally on other repatoms in their vicinity), whereas the retained continuum atoms are "local" (since their energy is computed using a local continuum approximation).

The repatoms constitute the nodes of an FE mesh as shown in Fig. 1b. Note that for computational convenience the mesh is continued into the atomistic region, although it serves no purpose there. The positions of discarded continuum atoms (appearing as open circles in Fig. 1b) are determined by FE interpolation:

$$\bar{\boldsymbol{q}}_i^{\text{c}} = \boldsymbol{Q}_i^{\text{c}} + \sum_{J=1}^{N_{\text{node}}} N_J(\boldsymbol{Q}_i^{\text{c}}) \, \boldsymbol{u}_J, \qquad (27)$$

where $\boldsymbol{Q}_i^{\text{c}}$ is the reference position of continuum atom $i$, $N_J$ is the FE shape function associated with node $J$, $\boldsymbol{u}_J$ is the displacement of node $J$, and $N_{\text{node}}$ is the number of nodes. Thus, the total number of degrees of freedom is reduced from $3N$ to $3N_{\text{node}}$.

Using the FE discretization, the sum over continuum atoms in Eq. (24) is approximated by employing the Cauchy-Born rule, which maps the continuum deformation to the motion of the underlying continuum atoms (Tadmor and Miller 2011):

$$\sum_{i=1}^{N^{\mathrm{c}}} \psi_i^{\mathrm{c}}(\boldsymbol{q}^{\mathrm{at}}, \bar{\boldsymbol{q}}^{\mathrm{c}}; \beta) \approx \sum_{e=1}^{N_{\mathrm{elem}}} \nu_e\, \psi^{\mathrm{CB}}(\boldsymbol{F}_e; \beta). \tag{28}$$

Here $\nu_e$ is the number of continuum atoms associated with element $e$ that fall within the continuum region, $\boldsymbol{F}_e$ is the deformation gradient of element $e$ (computed from the FE interpolation), and $\psi^{\mathrm{CB}}(\boldsymbol{F}_e; \beta)$ is the potential part of the Helmholtz free energy per atom computed for an infinite crystal deformed by $\boldsymbol{F}_e$ at inverse temperature $\beta$. (Note that for multilattice crystals, the displacement of basis atoms must be accounted for; see Tadmor et al. 1999.)

Putting everything together, the QC effective potential is given by

$$\mathcal{V}^{\mathrm{QC}}(\boldsymbol{q}^{\mathrm{at}}; \beta) = \min_{\boldsymbol{u}^{\mathrm{c}}} \left\{ \sum_{i=1}^{N^{\mathrm{at}}} V_i^{\mathrm{at}}(\boldsymbol{q}^{\mathrm{at}}, \boldsymbol{u}^{\mathrm{c}}) + \sum_{e=1}^{N_{\mathrm{elem}}} \nu_e\, \psi^{\mathrm{CB}}(\boldsymbol{F}_e; \beta) \right\}. \tag{29}$$

The hot-QC Hamiltonian is then

$$\mathcal{H}^{\mathrm{QC}}(\boldsymbol{q}^{\mathrm{at}}, \boldsymbol{p}^{\mathrm{at}}; \beta) = \mathcal{V}^{\mathrm{QC}}(\boldsymbol{q}^{\mathrm{at}}; \beta) + \mathcal{K}^{\mathrm{at}}(\boldsymbol{p}^{\mathrm{at}}; \beta). \tag{30}$$

A hot-QC simulation corresponds to an MD simulation of the atoms in the atomistic region using the QC Hamiltonian in Eq. (30).

The hot-QC formulation derived so far involves the minimization in Eq. (29) at every time step as $\boldsymbol{q}^{\mathrm{at}}$ is updated. A further approximation referred to as "hot-QC-dynamic" allows both nonlocal atoms and continuum nodes to evolve simultaneously by adding a correction term to account for artificial entropy associated with the mesh. With this correction, second-order accuracy in $k_{\mathrm{B}}T$ is retained. See Dupuy et al. (2005) and Tadmor et al. (2013) for details.

## 4    Integrating Temporal Acceleration into Spatial Coarse Graining

The hot-QC approach presented in the previous section significantly reduces the number of degrees of freedom in a system, thereby reducing the computational cost, but remains limited by the very short time step ($\sim 1\,\mathrm{fs}$) required for integrating the equations of motion of the nonlocal repatoms (i.e., atoms in the atomistic region). This severely limits the duration of hot-QC simulations. As a result, the loading rates in hot-QC simulations (as in MD simulations) are typically far larger than in experiments (otherwise no appreciable deformation would occur). For example, the hot-QC simulations of nanoindentation reported in Dupuy et al. (2005) were performed with an indenter velocity of about $5 \times 10^6\,\mu\mathrm{m/s}$, whereas the experimental loading rate is typically in the range of $0.001$–$1\,\mu\mathrm{m/s}$. Such large differences in rates can strongly influence the response of material systems, even leading to qualitative changes in some cases. Methods are therefore needed to accelerate the dynamics

occurring on the molecular scale. One approach suited for solid systems takes advantage of the separation between atomic vibrations and the far larger time scales associated with the nucleation and motion of defects. The idea is to skip over the uninteresting vibrations directly to the more infrequent defect-related events that evolve the system. Below is a description of this approach in a fully atomistic setting, followed by a discussion of how it can be incorporated into hot-QC.

## 4.1   Infrequent Events and Transition State Theory

The energetics of a system of $N$ particles can be represented in terms of a potential energy surface (PES). The PES is a conceptual plot of the energy of the system as a function of the positions of the atoms. Think of it as a topographical map in $3N + 1$ dimensions. The PES includes energy minima that are associated with equilibrium states of the system, separated from each other by dividing surfaces. The low points on the dividing surfaces are saddle points through which the system can most easily cross from one minimum to another; these are the transition states of the system. A two-dimensional schematic of a PES is shown in Fig. 2.

A solid system at a sufficiently low temperature spends most of its time vibrating within the basins of attraction of equilibrium states with occasional, rapid, thermally activated transitions between states. The trajectory of a typical system is represented as the black line in Fig. 2. The transition rate $R_{A \to B}$ is defined as the frequency of transitions from state A to state B (i.e., the number of transitions from A to B per
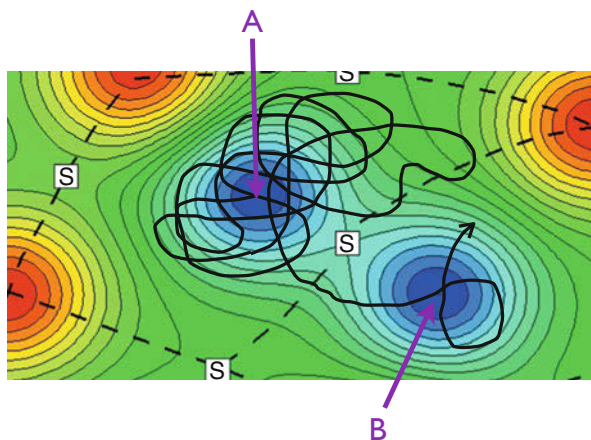


**Fig. 2** A PES for a system with two degrees of freedom represented by the horizontal and vertical axes. The colors represent the energy of the system with blue being lowest and red highest. Two equilibrium states (minima) are shown, denoted as A and B. The dashed lines are the dividing surfaces separating the region associated with one equilibrium state ("basin of attraction") from surrounding states. The low points on the dividing surfaces are saddle points denoted by S. The black line represents a trajectory of the system

unit time). It is important to understand that this rate is defined in the thermodynamic limit where the system spends an infinite amount of time transitioning back and forth between A and B. It is given by

$$R_{A \to B} \equiv \lim_{t \to \infty} \frac{N_{A \to B}(t)}{t_A(t)}, \tag{31}$$

where $N_{A \to B}(t)$ is the number of crossings from A to B during time $t$ and $t_A(t)$ is the portion of time spent in state A during time $t$. Inverting Eq. (31) gives the average waiting time that the system spends in state A before escaping:

$$\bar{t}_A^{\text{wait}} \equiv \lim_{t \to \infty} \frac{t_A(t)}{N_{A \to}(t)} = \frac{1}{R_{A \to}}. \tag{32}$$

This expression has been generalized to the case where there can be multiple escapes out of A. The total number of escapes out of state A during time $t$ is $N_{A \to}(t)$.

Under the assumption that every crossing of the dividing surface corresponds to a transition between states, the transition rate in Eq. (31) can be computed within the TST formalism of statistical mechanics (see, e.g., Vanden-Eijnden and Tal 2005). This is an approximation since dynamical trajectories that cross the dividing surface may return back to their original state instead of equilibrating in the new state (see Fig. 2 where the trajectory briefly crosses the dividing surface twice before making a true transition to state B). Neglecting these dynamical recrossings, the TST expression for Eq. (31) is

$$R_{A \to B}^{\text{TST}} = \frac{1}{2} \frac{\int_{\Sigma_{AB}} d\boldsymbol{q} \int d\boldsymbol{p} \ |v_n| \, e^{-\beta \mathcal{H}(\boldsymbol{q}, \boldsymbol{p})}}{\int_{\Omega_A} d\boldsymbol{q} \int d\boldsymbol{p} \ e^{-\beta \mathcal{H}(\boldsymbol{q}, \boldsymbol{p})}} = \langle |v_n| \, \delta_{AB}(\boldsymbol{q}) \rangle_A, \tag{33}$$

where $\Sigma_{AB}$ is a portion of the dividing surface separating A and B, $\Omega_A$ is a region in the configuration space (basin of attraction) associated with A, and $v_n$ is the velocity normal to the dividing surface. In the second expression, $\delta_{AB}$ is the Dirac delta, which is zero everywhere except on the dividing surface between A and B, and the subscript A indicates that the phase average is computed over $\Omega_A$. The TST rate defined in Eq. (33) is an equilibrium property of the system.

At low temperatures, the TST rate can be evaluated analytically by expanding the potential energy in the numerator about the saddle point configuration S and the potential energy in the denominator about A and applying the harmonic approximation (Sect. 2.3). The result is the famous Arrhenius equation that governs many physical processes:

$$R_{A \to B}^{\text{hTST}} = \nu e^{-\beta \mathcal{E}_a}, \tag{34}$$

where $\nu$ is an attempt frequency related to the curvatures at states A and S and $\mathcal{E}_a = \mathcal{V}(\boldsymbol{q}_S) - \mathcal{V}(\boldsymbol{q}_A)$ is the activation energy (the energy barrier for crossing from A to B across S).

## 4.2    Transitions in Hot-QC

As explained in Sect. 3.1, hot-QC is designed to reproduce equilibrium phase averages. As a result hot-QC simulations should exhibit transition behavior that is consistent with TST (since the TST rate itself is an equilibrium property as noted in Sect. 4.1). This is indeed correct provided that two conditions are satisfied (Kim et al. 2014):

1. All atoms significantly affected by the transition are inside the atomistic region.
2. The dividing surface can be approximated as a function of only atoms in the atomistic region.

Both assumptions are consistent with the basic hot-QC ansatz in Eq. (19) of limiting consideration to phase averages over degrees of freedom in the atomistic region. Under these conditions, it can be shown that

$$
\left( R_{\mathsf{A}\to\mathsf{B}}^{\mathrm{TST}} \right)_{\mathcal{V}} = \frac{1}{2} \frac{\int_{\Sigma_{\mathsf{AB}}(\boldsymbol{q}^{\mathrm{at}})} d\boldsymbol{q}^{\mathrm{at}} \int d\boldsymbol{p} \, |v_{\mathrm{n}}| \, e^{-\beta \mathcal{H}(\boldsymbol{q},\boldsymbol{p})}}{\int_{\Omega_A(\boldsymbol{q}^{\mathrm{at}})} d\boldsymbol{q}^{\mathrm{at}} \int d\boldsymbol{p} \, e^{-\beta \mathcal{H}(\boldsymbol{q},\boldsymbol{p})}}
$$

$$
= \frac{1}{2} \frac{\int_{\Sigma_{\mathsf{AB}}(\boldsymbol{q}^{\mathrm{at}})} d\boldsymbol{q}^{\mathrm{at}} \int d\boldsymbol{p}^{\mathrm{at}} \, |v_{\mathrm{n}}^{\mathrm{at}}| \, e^{-\beta \mathcal{H}^{\mathrm{QC}}(\boldsymbol{q}^{\mathrm{at}},\boldsymbol{p}^{\mathrm{at}})}}{\int_{\Omega_A(\boldsymbol{q}^{\mathrm{at}})} d\boldsymbol{q}^{\mathrm{at}} \int d\boldsymbol{p} \, e^{-\beta \mathcal{H}^{\mathrm{QC}}(\boldsymbol{q}^{\mathrm{at}},\boldsymbol{p}^{\mathrm{at}})}} = \left( R_{\mathsf{A}\to\mathsf{B}}^{\mathrm{TST}} \right)^{\mathrm{QC}}. \qquad (35)
$$

Thus TST rates in the QC system equal those in the fully atomistic system for transitions occurring in the atomistic region. This analytical result was verified numerically by Kim et al. (2014).

## 4.3    Accelerated Molecular Dynamics (Hyperdynamics)

Equation (34) provides guidance on how to increase transition rates in systems governed by TST; one can either decrease the activation energy $\mathcal{E}_{\mathrm{a}}$ or increase the temperature (i.e., reduce $\beta$). These two approaches are adopted in hyperdynamics (Voter 1997) and temperature-accelerated dynamics (Sørensen and Voter 2000), respectively. Here we focus on hyperdynamics.

Consider a system with positions $\boldsymbol{q}$ that is within the basin of attraction $\Omega_{\mathsf{A}}$ of state $\mathsf{A}$. To expedite the escape from this state, the true potential energy of the system $\mathcal{V}(\boldsymbol{q})$ is modified by the addition of a nonnegative "bias potential" $\Delta \mathcal{V}(\boldsymbol{q})$, such that

$$
\mathcal{V}_{\mathrm{b}}(\boldsymbol{q}) = \mathcal{V}(\boldsymbol{q}) + \Delta \mathcal{V}(\boldsymbol{q}), \qquad (36)
$$

in which

$$
\Delta \mathcal{V}(\boldsymbol{q}) = \begin{cases} > 0 & \text{if } \boldsymbol{q} \in \Omega_{\mathsf{A}}, \\ = 0 & \text{if } \boldsymbol{q} \in \Sigma_{\mathsf{A}}, \end{cases} \qquad (37)
$$

where $\Sigma_A$ is the dividing surface separating state A from surrounding states. (Methods for constructing bias potentials are discussed below.)

An MD simulation using $\mathcal{V}_b(\boldsymbol{q})$ in Eq. (36) will by design escape from state A more quickly than the original system with potential $\mathcal{V}(\boldsymbol{q})$. However, since the biased system has a different potential energy, it is in fact a different material. Why then would the results obtained with $\mathcal{V}_b$ be relevant to the original material? The key lies in the condition in Eq. (37) that requires $\Delta\mathcal{V}$ to vanish on the dividing surfaces. It is shown next that if this is satisfied, then the state-to-state dynamics of the original and biased systems will be the same. In other words, if both systems were simulated for a very long time, they would visit each state with the same relative probability, even though the dynamics within each state would be different. All of this holds only within the TST formalism (Sect. 4.1); thus the state-to-state dynamics of the biased and original systems *would* differ due to dynamic recrossings, an effect that becomes more pronounced with increasing temperature (Vanden-Eijnden and Tal 2005).

In order to prove the state-to-state dynamics assertion, recall the TST rate expression in Eq. (33). Separating the Hamiltonian into kinetic and potential contributions and multiplying the denominator and numerator by $\exp(+\beta\Delta\mathcal{V})\exp(-\beta\Delta\mathcal{V}) = 1$, the following relation is obtained:

$$R_{A\to B}^{TST} = \frac{1}{2}\frac{\int_{\Sigma_{AB}} d\boldsymbol{q}\int d\boldsymbol{p}\ |v_n|\ e^{-\beta(\mathcal{K}+\mathcal{V}+\Delta\mathcal{V})}e^{\beta\Delta\mathcal{V}}}{\int_{\Omega_A} d\boldsymbol{q}\int d\boldsymbol{p}\ e^{-\beta(\mathcal{K}+\mathcal{V}+\Delta\mathcal{V})}e^{\beta\Delta\mathcal{V}}} = \frac{\langle|v_n|\,\delta_{AB}(\boldsymbol{q})\rangle_{A_b}}{\langle e^{\beta\Delta\mathcal{V}}\rangle_{A_b}}. \qquad (38)$$

In the second expression, the subscript $A_b$ indicates a phase average on the biased PES. This expression follows from the first by dividing the numerator and denominator by the biased partition function restricted to state A:

$$\int_{\Omega_A} d\boldsymbol{q}\int d\boldsymbol{p}\ e^{-\beta(\mathcal{K}+\mathcal{V}+\Delta\mathcal{V})}.$$

Next, enforcing the condition in Eq. (37) that $\Delta\mathcal{V} = 0$ on $\Sigma_{AB}$ and rearranging, Eq. (38) becomes

$$R_{A_b\to B}^{TST} = \left\langle e^{\beta\Delta\mathcal{V}}\right\rangle_{A_b} R_{A\to B}^{TST}, \qquad (39)$$

where the definition in Eq. (33) was used. This result shows that the escape rate from the biased state $A_b$ is boosted by the exponential factor $\left\langle e^{\beta\Delta\mathcal{V}}\right\rangle_{A_b}$. (Since $\Delta\mathcal{V} \geq 0$, the boost is always greater or equal to one.) Further, considering two possible escapes, $A \to B$ and $A \to C$, the relative probabilities are preserved on the biased PES since

$$\frac{R_{A_b\to B}^{TST}}{R_{A_b\to C}^{TST}} = \frac{\left\langle e^{\beta\Delta\mathcal{V}}\right\rangle_{A_b} R_{A\to B}^{TST}}{\left\langle e^{\beta\Delta\mathcal{V}}\right\rangle_{A_b} R_{A\to C}^{TST}} = \frac{R_{A\to B}^{TST}}{R_{A\to C}^{TST}} \qquad (40)$$

This is the basis for the statement that state-to-state dynamics are preserved by the biased system (Voter 1997).

The remaining step is to relate the time scale of the biased system dynamics to the original system. For example, if the biased system escapes from a basin in 1 ns, what does this correspond to in the original system? This is a subtle question. To understand this, recall the TST definition of the average waiting time in Eq. (32). Using Eq. (39) and rearranging, the true waiting time is related to the biased waiting time by

$$\bar{t}_A^{wait} = \left\langle e^{\beta \Delta \mathcal{V}} \right\rangle_{A_b} \bar{t}_{A_b}^{wait}. \tag{41}$$

Converting from a phase average to the time average performed in an MD simulation, one can relate the MD time step of the biased system $\Delta t_{MD}$ to the corresponding boosted time step for the real system:

$$\Delta t_{boosted}^k = e^{\beta \Delta \mathcal{V}(\boldsymbol{q}^k)} \Delta t_{MD}, \tag{42}$$

where $k$ is the time step index and $\boldsymbol{q}^k$ is the corresponding atomic configuration. The stepwise definition in Eq. (42) will give statistics that are consistent with Eq. (41) (Voter 1997).

The key step in setting up and running a hyper-QC simulation is the construction of a bias potential that is zero on dividing surfaces (as stated in Eq. (37)) and provides good boost. In the original hyperdynamics method (Voter 1997), the lowest eigenvalues and corresponding eigenvectors of the Hessian were used to construct the bias potential. However, this approach can be computationally expensive. Later, a bias potential without significant computational overhead, called the "bond-boost method," was proposed by Miron and Fichthorn (2003). The bond-boost method is based on the increased changes in bond lengths that occur when a system makes transitions. In particular, a bias potential $\Delta \hat{\mathcal{V}}(e_n)$ can be defined in terms of the fractional bond length change, $e_n = |l_n - l_n^{ref}| / l_n^{ref}$, where $l_n$ is the bond length of the $n$th bond and $l_n^{ref}$ is the bond length in the reference configuration (usually taken as the potential energy minimum). The functional form $\Delta \hat{\mathcal{V}}(e_n)$ is selected to guarantee that if any $e_n$ exceeds a threshold value (which can be empirically estimated by preprocessing simulations), the total bias potential vanishes. Other bias potentials have been proposed including one which uses the hyperdistance from a reference configuration (Kim and Falk 2010) and a "mechanism-based bias potential" that makes use of the specific physics being modeled to obtain better boosts (Kim and Tadmor 2017).

## 4.4 Temporal Acceleration in Hot-QC (Hyper-QC)

The last step is to integrate hyperdynamics within the hot-QC framework to create the hyper-QC method. Adopting the conditions outlined in Sect. 4.2, it is straightforward to show that the hyperdynamics derivation carries over to hot-QC

with a bias potential that is a function of only atoms retained in the atomistic region (Kim et al. 2014):

$$\mathcal{V}_b^{QC}(\boldsymbol{q}^{at}; \beta) = \mathcal{V}^{QC}(\boldsymbol{q}^{at}; \beta) + \Delta\mathcal{V}^{QC}(\boldsymbol{q}^{at}; \beta). \qquad (43)$$

A hyper-QC simulation corresponds to doing MD with $\mathcal{V}_b^{QC}$. Analogous to standard hyperdynamics, hyper-QC preserves the original state-to-state dynamics of hot-QC under the TST assumptions (Kim et al. 2014). See Kim and Tadmor (2017) for a detailed discussion of the practical issues in setting up and performing hyper-QC simulations.

Hyper-QC was first applied to a one-dimensional system to verify the formulation (Kim et al. 2014). A chain of 1000 atoms undergoing a transition from a bonded state to an unbonded state at constant temperature was studied. The hyper-QC simulations exactly recovered the TST rates of the fully atomistic model, running 77 to 92 times faster in terms of computation time.

A hyper-QC application to nanoindentation of a face-centered cubic (fcc) nickel crystal is described in Kim and Tadmor (2017). The single crystal is of size $2000 \times 100$ Å with periodic boundary conditions corresponding to the minimal repeated distance of 4.31 Å applied in the out-of-plane direction. The indenter is cylindrical with a diameter of 100 Å. The hyper-QC model contains 6,135 repatoms, which represent a fully atomistic system of 79,461 atoms. The atomic interactions are modeled by an embedded atom method (EAM) potential (Zhou et al. 2004). A mechanism-based bias potential designed to accelerate slip events in fcc is employed.

Simulations were performed at three temperatures, $T = 100, 200$, and $300$ K, and indenter velocities ranging from $10^6$ to $10^2$ μm/s. Results for the critical force required to nucleate the first dislocation under the indenter are plotted in Fig. 3. A typical logarithmic dependence on the indenter velocity is observed, indicating that the dislocation nucleation is thermally activated. Based on this observation,

**Fig. 3** The indentation force at which the first partial dislocation nucleates as a function of indenter velocity at temperatures of 100, 200, 300 K. The straight lines are obtained from the Tomlinson-Prandtl model. (Reprinted with permission from Kim and Tadmor 2017)
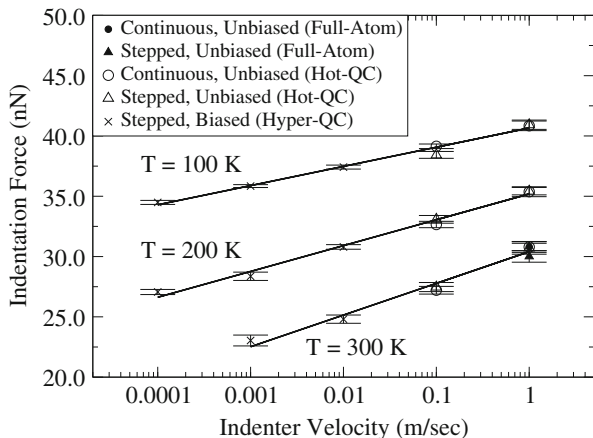
a simple Prandtl-Tomlinson model is constructed (Tomlinson 1929; Gnecco et al. 2000; Tadmor et al. 2013). The predictions of the model are shown as the straight lines in Fig. 3. (In Fig. 3, the model was fit separately at each temperature, and a fit over all temperatures shows more scatter but is still good overall; see Kim and Tadmor 2017.)

The smallest indentation rate of $10^2 \, \mu$m/s achieved in the hyper-QC simulation is significantly lower than that attainable using hot-QC and is approaching experimental loading rates (see start of Sect. 4). This is possible due to speedups in hyper-QC ranging from 1000 (at $T = 300$ K) to 10,000 (for $T = 100$ and 200 K) compared with hot-QC. This speedup includes both the spatial coarse graining and temporal acceleration.

## 4.5   Other Approaches: Finite-Temperature CQC

The extension of QC to finite temperature, dynamics, and temporal acceleration described above (i.e., hot-QC and hyper-QC) is based on equilibrium statistical mechanics. A different approach is adopted in the CQC method. Based on the maximum entropy principle of Jaynes (1957a, b) and a mean field approximation (Kulkarni et al. 2008; Venturini et al. 2014), the finite-temperature CQC method seeks a probability distribution $\rho(\mathbf{q}, \mathbf{p})$ of the form:

$$\rho \propto \exp\left[ -\sum_{i=1}^{N} \beta_i \left\{ \frac{m_i \omega_i^2 ||\mathbf{q}_i - \bar{\mathbf{q}}_i||^2}{2} + \frac{||\mathbf{p}_i - \bar{\mathbf{p}}_i||^2}{2m_i} \right\} \right], \tag{44}$$

which maximizes the entropy functional $S$ defined as

$$\mathcal{S}[\rho] \equiv -k_{\mathrm{B}} \langle \ln \rho \rangle = -k_{\mathrm{B}} \int \rho \ln \rho \, dq \, dp. \tag{45}$$

The maximum is sought subject to local constraints, $\langle h_i \rangle = e_i$ $(i = 1, \ldots, N)$, where $e_i$ is regarded as the energy of atom $i$ observed on a macroscopic time scale. In Eq. (44), $\beta_i$, $\bar{\mathbf{q}}_i$, and $\bar{\mathbf{p}}_i$ are the local temperature, mean position, and mean momentum of atom $i$, respectively, and $\boldsymbol{\omega}$ are free parameters to be determined.

For a quasistatic process, $\bar{\mathbf{q}}$, $\bar{\mathbf{p}}$, and $\boldsymbol{\omega}$ are determined by maximizing the canonical free entropy $\Phi$ given by (Venturini et al. 2014)

$$\Phi(\boldsymbol{\beta}; \bar{\mathbf{q}}, \bar{\mathbf{p}}, \boldsymbol{\omega}) = -k_{\mathrm{B}} \sum_{i=1}^{N} \left[ \beta_i \left( \frac{||\bar{\mathbf{p}}_i||^2}{2m_i} + \langle V_i \rangle \right) - \frac{3}{2} + 3 \ln \left( \frac{\beta_i \omega_i}{2\pi} \right) \right]. \tag{46}$$

A Hamiltonian dynamics is postulated for the evolution of the mean positions and momenta $(\bar{\mathbf{q}}, \bar{\mathbf{p}})$ (Ponga et al. 2015, 2016):

$$\dot{\bar{\mathbf{q}}}_i = \frac{\partial \mathcal{H}_{\mathrm{ME}}}{\partial \bar{\mathbf{p}}_i}, \qquad \dot{\bar{\mathbf{p}}}_i = -\frac{\partial \mathcal{H}_{\mathrm{ME}}}{\partial \bar{\mathbf{q}}_i}, \tag{47}$$

where $\mathcal{H}_{\text{ME}}$ is the max-ent Hamiltonian:

$$\mathcal{H}_{\text{ME}} = -\sum_{i=1}^{N} \frac{1}{k_{\text{B}}} \frac{\partial \Phi}{\partial \beta_i}. \tag{48}$$

Next, an FE-based coarse-graining scheme is introduced to reduce the number of degrees of freedom using the fully nonlocal cluster-based summation method of Knap and Ortiz (2001).

In contrast to hot-QC, which is based on equilibrium statistical mechanics and assumes a system at constant uniform temperature, the finite-temperature CQC method can be applied to nonequilibrium systems with nonuniform temperature fields. As such, Eqs. (46) and (47) can be coupled with a phenomenological atomic-level heat conduction equation to determine $\beta$ (Venturini et al. 2014). Moreover, since the equations of motion are for the *mean* positions and momenta, where the short-time thermal vibrational modes are averaged out, the resultant trajectory will be smooth on microscopic time scales enabling much larger time steps than those used in conventional MD simulations (Ponga et al. 2015).

## 5    Concluding Remarks

The size of systems that can be simulated using fully atomistic MD continues to increase as computing power grows, with the current world record well above one trillion atoms (Germann and Kadau 2008). Given this, one may wonder regarding the need for accelerated continuum-atomistic methods that aim to reduce degrees of freedom and extend time scales in molecular simulations. There are several reasons why such methods remain of interest. First, as more complex materials are modeled, the computational expense of interatomic models greatly increases, dramatically reducing the size of systems that can be studied. This is the case for state-of-the-art reactive force fields, bond-order potentials, and machine learning-based potentials that are of current interest (Tadmor and Miller 2011). Second, without acceleration techniques, the time scales attainable in molecular simulations are typically very short leading to loading rates that are many orders of magnitude higher than experiments except under extreme conditions. Third, the reduction in computation time promised by the methods described in this article will make it possible to perform multiple simulations of a given system and thereby obtain uncertainty estimates in terms of the various physical and numerical parameters in the problem. Finally, there is the philosophical view that expending huge amounts of effort and energy to simulate the dynamics of trillions of atoms that are mostly obeying continuum field theories is not the best use of resources. The challenge lies in developing robust and predictive multiscale methods that are able to span multiple length and time scales and are applicable to a broad range of materials under both equilibrium and nonequilibrium conditions. These topics remain an exciting and ongoing area of research.

# References

Curtin WA, Miller RE (2003) Atomistic/continuum coupling methods in multi-scale materials modeling. Model Simul Mater Sci Eng 11:R33–R68

Dupuy LM, Tadmor EB, Miller RE, Phillips R (2005) Finite temperature quasicontinuum: molecular dynamics without all the atoms. Phys Rev Lett 95:060202

Germann TC, Kadau K (2008) Trillion-atom molecular dynamics becomes a reality. Int J Mod Phys C 19:1315–1319

Gnecco E, Bennewitz R, Gyalog T, Loppacher C, Bammerlin M, Meyer E, Güntherodt HJ (2000) Velocity dependence of atomic friction. Phys Rev Lett 84:1172–1175

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115:9657–9666

Jaynes ET (1957a) Information theory and statistical mechanics. Part I. Phys Rev 106:620–630

Jaynes ET (1957b) Information theory and statistical mechanics. Part II. Phys Rev 108:171–190

Kim WK, Falk ML (2010) Accelerated molecular dynamics simulation of low-velocity frictional sliding. Model Simul Mater Sci Eng 18:034003

Kim WK, Tadmor EB (2017) Accelerated quasicontinuum: a practical perspective on hyper-QC with application to nanoindentation. Philos Mag 97:2284–2316

Kim WK, Luskin M, Perez D, Voter AF, Tadmor EB (2014) Hyper-QC: an accelerated finite-temperature quasicontinuum method using hyperdynamics. J Mech Phys Solids 63:94–112

Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300–313

Knap J, Ortiz M (2001) An analysis of the quasicontinuum method. J Mech Phys Solids 49:1899–1923

Kulkarni Y, Knap J, Ortiz M (2008) A variational approach to coarse graining of equilibrium and non-equilibrium atomistic description at finite temperature. J Mech Phys Solids 56:1417–1449

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99:12562–12566

LeSar R, Najafabadi R, Srolovitz D (1989) Finite-temperature defect properties from free-energy minimization. Phys Rev Lett 63:624–627

Li J, Sarkar S, Cox WT, Lenosky TJ, Bitzek E, Wang Y (2011) Diffusive molecular dynamics and its application to nanoindentation and sintering. Phys Rev B 84:054103

Miller RE, Tadmor EB (2002) The quasicontinuum method: overview, applications, and current directions. J Comput Aided Mater Des 9:203–239

Miller RE, Tadmor EB (2009) A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Model Simul Mater Sci Eng 17:053001

Miron RA, Fichthorn KA (2003) Accelerated molecular dynamics with the bond-boost method. J Chem Phys 119:6210–6216

Ponga M, Ortiz M, Ariza M (2015) Finite-temperature non-equilibrium quasi-continuum analysis of nanovoid growth in copper at low and high strain rates. Mech Mater 90:253–267

Ponga M, Ramabathiran AA, Bhattacharya K, Ortiz M (2016) Dynamic behavior of nano-voids in magnesium under hydrostatic tensile stress. Model Simul Mater Sci Eng 24:065003

Shenoy VB, Miller R, Tadmor EB, Rodney D, Phillips R, Ortiz M (1999) An adaptive finite element approach to atomic-scale mechanics: the quasicontinuum method. J Mech Phys Solids 47:611–642

Sørensen MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 112:9599–9606

Tadmor EB, Miller RE (2011) Modeling materials: continuum, atomistic and multiscale techniques. Cambridge University Press, Cambridge

Tadmor EB, Ortiz M, Phillips R (1996) Quasicontinuum analysis of defects in solids. Philos Mag A 73:1529–1563

Tadmor EB, Smith GS, Bernstein N, Kaxiras E (1999) Mixed finite element and atomistic formulation for complex crystals. Phys Rev B 59:235–245

Tadmor EB, Legoll F, Kim WK, Dupuy LM, Miller RE (2013) Finite-temperature quasi-continuum. Appl Mech Rev 65:010803

Tomlinson GA (1929) A molecular theory of friction. Philos Mag 7:905–939

Vanden-Eijnden E, Tal FA (2005) Transition state theory: variational formulation, dynamical corrections, and error estimates. J Chem Phys 123:184103

Venturini G, Wang K, Romero I, Ariza MP, Ortiz M (2014) Atomistic long-term simulation of heat and mass transport. J Mech Phys Solids 73:242–268

Voter AF (1997) A method for accelerating the molecular dynamics simulation of infrequent events. J Chem Phys 106:4665–4667

Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57: 13985–13988

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, BP Uberagua (eds) Radiation effects in solids. Springer, NATO Publishing Unit, Dordrecht, pp. 1–23

Zhou XW, Johnson RA, Wadley HNG (2004) Misfit-energy-increasing dislocations in vapor-deposited CoFe/NiFe multilayers. Phys Rev B 69:144113

# Long-Timescale Simulations: Challenges, Pitfalls, Best Practices, for Development and Applications

# 38

Graeme Henkelman, Hannes Jónsson, Tony Lelièvre,
Normand Mousseau, and Arthur F. Voter

**Abstract**

In this chapter, we examine the practice of developing, implementing, and applying long-timescale simulation methods. In contrast to standard molecular dynamics, the performance, and sometimes the accuracy, of long-timescale atomistic methods can vary greatly from one application to another. Therefore, for the practitioners, it is particularly important to understand the strengths and weaknesses of the methods, in order to assess their respective potential for

G. Henkelman
Department of Chemistry and the Institute for Computational and Engineering Sciences,
University of Texas at Austin, Austin, TX, USA

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX, USA
e-mail: henkelman@utexas.edu

H. Jónsson
Faculty of Physical Sciences, University of Iceland, Reykjavík, Iceland

Department of Energy Conversion and Storage, Technical University of Denmark, Lyngby,
Denmark
e-mail: hj@hi.is

T. Lelièvre
CERMICS, École des Ponts, ParisTech, INRIA, Champs-sur-Marne, France
e-mail: tony.lelievre@enpc.fr

N. Mousseau
Departement de physique and Regroupement québécois sur les matériaux de pointe, Université de
Montréal, Montréal, QC, Canada

Department of Physics, University de Montréal, Montréal, Canada
e-mail: nm@normandmousseau.com

A. F. Voter
Theoretical Division T-1, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: afv@lanl.gov

specific problems, as well as maximize their efficiency. For the method developer, clearly assessing the challenges faced by current methods as well as the areas of opportunities for future development is paramount.

In the following, we present the opinion of five leaders in the field regarding best practices, challenges, and pitfalls in the use and development of such methods. Their answers both provide a roadmap to how best to approach the field and deliver insight into areas that need addressing in the future.

## Question 1: In your experience, what are the biggest challenges associated with extending the timescale of atomistic simulations? What pitfalls does one need to be aware of?

AFV: Biggest challenges: Low barriers, low barriers, low barriers! In general, the lowest barriers in a system may be substantially lower than the relevant higher barriers, those barriers one needs to surmount to reach the timescale of interest. In this case, although there may be good acceleration compared to direct molecular dynamics, the gap between the fast rates (low barriers) and slow rates (high barriers) might prevent observation of the desired very-long-time dynamics, because a huge number of the lower-barrier events need to take place before a high-barrier event occurs. Especially frustrating is when the lowest barriers are actually so low, relative to $k_B T$, that there is very little absolute acceleration.

The very worst case is "persistent low barriers," in which the system keeps visiting new states, and each of these new states has low-barrier pathways that inhibit acceleration. For a state that has been visited previously (especially if it was visited many times previously), accumulated information can be used to improve the performance on the revisit. This cannot be done very effectively on the first visit, however, so that for this type of system the acceleration will be very low.

For systems with a serious low-barrier problem, there is hope for being able to combine many states that are connected by low barriers into a single superbasin state. The quasi-static distribution (QSD) formalization of parallel-replica dynamics offers a framework for this kind of generalization. In some cases, we have been able to improve the boost substantially by using our understanding of the system to apply such an approach. So far, however, it has proved challenging to develop an automated approach to this problem, one that would work, e.g., for proteins.

NM: I totally agree with this problem. At one point, however, the low-barrier problem mixes with the configuration entropy problem. In the case of proteins, for example, many barriers have a significant entropic contribution, meaning that the transition state cannot be well described with a single point on the energy landscape.

In materials, where the problem is simpler, we have to consider two types of low barriers: those that do not evolve the system, which are generally referred to as flickers, and those that are an intrinsic part of the evolution. Distinguishing between those requires a deep knowledge of the problem. Moreover, it is not always possible to draw a clear line between the two.

For kinetic Monte Carlo (KMC), the main problem is that every time one leaves a basin, it becomes necessary to reconstruct it, which means a lot of effort for a relatively short time step. When barriers remain energy-activated, it is nevertheless

possible to see a solution where basins, incorporating multiple states separated by low barriers, are generated on a local basis and merged if needed. As long as those states are distinct, KMC is efficient. When this is not the case, however, one can get overloaded by the cost of rebuilding the basin.

AFV: Unfortunately, pitfalls also abound. For example, the shape of the bias potential in hyperdynamics is a very subtle problem. In hyperdynamics, one must take care to design a bias potential that is zero at all dividing surfaces, so that it does not "block" any pathways (i.e., slow them down). A problem is that there is typically not any obvious signature when a pathway is being blocked. In principle, one can check for a nonzero bias at a dividing surface for an escape mechanism that has been observed, but for a mechanism that has not been observed because the rate was slowed down too much, no such check can be performed – then the dynamics are just wrong.

Another pitfall is the high dimensionality of typical problems in materials. Our intuition can be easily misled by drawings in one or two dimensions when the real system is 3N dimensional. An example is the flat bias potential of Steiner, Genilloud, and Wilkins (1998), in which the potential energy is replaced with a flat potential for energies lower than some threshold, with this threshold chosen to be lower than the energy of the lowest saddle point bounding the state (picture a frozen lake in the mountains). In one dimension, or a few dimensions, this simple form of bias potential can indeed give good acceleration, and it is natural to think that this characteristic would persist to high-dimensional systems. However, it does not. Although it is a valid bias, because the bias is guaranteed to be turned off whenever the potential energy is higher than the lowest barrier, it is no longer an effective bias because the typical potential energy in a system with N atoms is roughly $3Nk_BT/2$, which for large N will be much higher than the threshold energy, so that the instantaneous boost is rarely turned on. This is difficult or impossible to draw in a one-dimensional diagram.

Another example of where our intuition can fail us in higher dimension is for a bias potential that is constructed from space-filling objects in a low-dimensional collective-variable space, as in an approach based on metadynamics, for example. Then, although the acceleration can remain large as N increases, the bias form itself may no longer be valid for hyperdynamics. This is because any reaction coordinate that is orthogonal to the collective variable space will be blocked. Thus, extreme care must be exercised.

NM: This problem is not limited to hyperdynamics. Methods such as ART nouveau and the dimer evolve the system through this high-dimension landscape to find saddle points. Yet, this search does not always work and, often, it is clear that the failure to find a saddle point is related to the structure of the energy landscape, a structure that is almost impossible to figure out given the high-dimensionality of the problem. Similarly, without a detailed knowledge of the energy landscape, it is not possible to ensure that these methods can find, even in principle, all connected saddle points.

HJ: In my opinion, the most basic pitfall is to assume some transition mechanism(s), some reaction coordinate, and base the time acceleration scheme on that assumption. This can lead to incorrect time evolution of the system. There is a

myriad of cases where the mechanism for atomic scale transitions turned out to be entirely different from what one might guess a priori. The simulation should show what the relevant transition mechanism is, and not rely on a preconceived notion of the mechanism. There are many schemes that are based on such preconceived notion, metadynamics being one of them that is used frequently. It is also important to realize that full free energy sampling using such an assumed reaction coordinate is not going to make up for a bad choice of the reaction coordinate. If the sampling is carried out in a subspace (e.g., hyperplane) that is not lined up with the transition state of the relevant mechanism, then no matter how much sampling is carried out, the deduced rate of the transition will be wrong and the simulated time evolution incorrect.

**Question 2: What are the main limitations of current methods and what needs to be done to address them?**

NM: There has been considerable development over the last two decades regarding accelerated methods. While they have opened up new regimes of physics that had been out of reach until now, significant challenges remain.

(I) Cost-effective accurate force fields. Long-time simulations in materials typically require relatively large systems, as following kinetics often implies displacement. Ab initio calculations being still limited to boxes of 1000 atoms or less, they cannot be used directly in these simulations. For lattice-based atomistic KMC methods, it is possible to construct an event catalog using small cells and density functional theory (DFT). Yet, for complex systems with significant deformations or a large number of configurations, even catalog building becomes too costly with ab initio. When dealing with elemental systems, especially pure metals, empirical potentials can offer a decent level of precision. These potentials cannot be relied upon, however, for conformations far away from the close-packed states, in the presence of many elements or when considering semiconductors, magnetic elements, etc. Over the last decade, statistical-derived potentials, using neural networks or other automatic learning methods, have shown significant promise. There is still work to be done, however, before these methods can be used regularly. Moreover, these new approaches remain very costly, limiting their application.

(II) Entropic effects. KMC methods, whether off-lattice or lattice-based, are effectively run at zero K. Most of the time, entropic contributions are included through a constant prefactor that supposes that the local environments remain relatively similar throughout the simulation. Some groups go beyond this simple approximation and evaluate the local entropic contributions for each environment using the harmonic approximation of transition state theory (TST), which supposes a temperature-independent prefactor, an approximation also made in temperature accelerated dynamics (TAD), one of the accelerated MD (AMD) approaches. Yet, this is not always valid. For example, an atom moving in a three-vacancy cluster in an FCC metal forms a tetrahedral vacancy cluster centered around an atom, a structure stable at high temperature. As it turns out, this structure is very unstable at low temperature, although it exists. In Ni, this state is at an energy 0.4 eV above ground state, separated

from it by a barrier of only 0.08 eV. Its stabilization at high temperature is clearly due to an increase in thermal vibrations that prevent the atom from moving back into its original position, a feature that might not be captured by the harmonic approximation (S. Mahmoud, et al., Acta Materialia (in press)). It is therefore likely necessary to go beyond the harmonic approximation with thermodynamic integration methods that can be automated.

(III) Going beyond TST. Most accelerated methods rely on TST at some level. There is a need to work on this theory and see how one can expand it to complex systems without having to use the heavy tools developed in systems dominated by entropy, such as transition path sampling.

(IV) Handling flickers and energy barrier distributions efficiently. In spite of considerable advances over the last years, handling systems with a wide distribution of relevant barriers remains a challenge. Yet, following the evolution of a grain boundary, the formation of a nanostructure, or the aging of a multicomponent glass, involves working with continuous energy distributions.

(V) A more efficient cataloging. Recognizing local environments, whether using KMC or MD-based methods is an essential part of recycling previous efforts. Over the years, a number of approaches have been proposed – geometrical, lattice-based, and topological. Yet, all of them suffer from some limitations that decrease the recycling. There is a need to carefully study this aspect and identify methods that are flexible, can be applied to wide range of environments, provide useful comparison, and facilitate the reconstruction.

HJ: Current methods are mainly based on the identification of local minima on the potential energy surface (with the exception of parallel replica dynamics, though even there, in practice, states are often defined in terms of local energy minima). There is a large set of problems, where entropy plays a central role and energy barriers are small and numerous. The definition of states needs to be more in terms of subspaces where the system spends enough time to locally equilibrate and for which the time evolution can be described as a Markov chain. Full TST, as opposed to the harmonic approximation to TST, coupled with the variational principle can, I believe, deal with these kinds of systems, the problem is coming up with an efficient algorithm for representing and optimizing the dividing surface that defines a state. Very little work has been done on this so far. Is it timely now to give it a try?

TL: I think that in some sense, AMD methods can be seen as a way to go beyond the TST, especially parallel replica dynamics or parallel trajectory splicing. Indeed, the decorrelation step is a way to check if first order kinetics can be applied to model the exit event, and then, these algorithms do not require knowledge of the exit rates. However, the efficiency in parallel replica is limited by the decorrelation time within the state. Parallel trajectory splicing is a nice idea to overcome this problem, but we need to think more about algorithms that are able to exploit massively parallel machines.

AFV: Normand, Tony, and Hannes make good points. As discussed under question 1, in my mind the main limitation of the methods is how hard it is to get

good acceleration for systems with persistent low barriers. Although some progress has been made on this front by various groups, I have been surprised by how difficult this challenge has remained. Ever greater cleverness will be required.

Amplifying Normand's point on the difficulty in using electronic structure forces, although in principle AMD methods can be used as easily with an expensive first-principles force call as it can with an empirical potential, in practice this may be far from the truth. For electronic structure forces, it may take so much computer time to advance the dynamics that the system will still be in its initial state when the computer budget runs out. In this situation, while formally the boost factor may be high, if the system has not yet jumped to a new state, the effective boost factor is zero.

**Question 3: Are there specific issues or challenges associated with applying these methods as opposed to their development? What could be done to further their widespread adoption?**

GH: Echoing Normand's point above, perhaps the most important challenge associated with a more widespread application of accelerated methods is that they cannot, at present, be used with standard DFT. There is an enormous community of scientists using DFT to model dynamics in chemical and material science applications, but we have not provided accelerated methods which are sufficiently efficient to be used routinely with energies and forces from DFT. While current methods can be based upon empirical potentials, for example as implemented in parallel codes such as LAMMPS, there are a vanishingly small number of applications for which we have accurate potentials as compared to those which can be described by DFT.

Some of the problem, I believe, can be attributed to the electronic structure community, which puts a greater emphasis on the accuracy of each energy calculation rather than an accurate sampling of configuration space. Recent developments of surrogate models, including those based upon machine learning, have the potential to change this bias if sampling could be done at a fraction of the cost of DFT. Additionally, there is a potential application for more approximate methods which can be used directly with DFT. Regardless of which direction the connection is made, either more approximation and efficient methods or more approximate and efficient potentials, providing tools in a form which can be used directly by the community of people running DFT calculations is key to their widespread adoption.

TL: I have the feeling that one of the major difficulties when applying AMD techniques in a general setting is the definition of good metastable states. It would be great to be able to define automatically or adaptively (as the simulation runs) good metastable states. These metastable states should be such that:

(I) The time to leave these states is much larger than the time to reach local equilibrium (quasi-stationary distribution QSD), for a generic initial condition obtained when entering the state.

(II) The discrepancy between the transition rates associated with these states is not too large (otherwise the algorithm spends much of the time switching over low barriers).

(III)  The states give a reasonable description of the macroscopic state of the system (since the details of the dynamics within states are lost).
(IV)  There is a way to estimate the convergence time to local equilibrium faithfully.

The list is rather long, but one could think of using modern statistical techniques to extract from the trajectories "good" states. The fact that these states do not need to define a partition of the state space could be used to make the construction easier. This is very much related to discussions about choosing good reaction coordinated or collective variables, which is obviously a difficult question, but the interest of AMD methods is that even if some of the states are not really well chosen, the methods can still give reliable results, e.g., with parallel replica dynamics.

AFV: I have been somewhat surprised that effective application of the AMD methods seems to require serious dedication on the part of the user. In essence, the user must become an expert; this takes time, and some users become discouraged before reaching this stage. Further automation of the methods, as discussed above, should help, and we may achieve this in the future, so that for a nonexpert user, applying an AMD code could be a "turnkey" operation.

NM: In addition to the challenges already mentioned, with which I totally agree, I would add that the current accelerated methods, based on high energy barriers with respect to $k_B T$, are restricted to solids well below melting. This is a considerable limitation as it prevents us from looking at molecules in solvent, including biomolecules, many catalytic and growth processes and a number of other fundamental questions. Overcoming this limitation is not impossible but it will require rethinking the approach and, more important, reworking TST.

HJ: Transition state theory actually works fine in cases where the free energy barrier is mainly of entropic origin. The effusion through a hole in a cavity is a nice example of how TST can give exact results even where there is no energy barrier. What is missing, however, is an efficient implementation of full TST where the dividing surface is systematically optimized (using the variational principle) to obtain the free energy barrier and thereby the mechanism(s) of the transition. But, I want to emphasize that I agree with all that has been written here above.

**Question 4: In developing or applying an approach, what are the best practices you recommend?**

HJ: The first rule is to have one or more test problems that are simple enough or well enough established so that the answer that should be obtained from the calculation is known. For example, when testing a method for estimating a transition rate, choose a system where the energy barrier is low enough and the temperature high enough that relatively long, but not too long, simulations of the dynamics using basic equations of motion can be used for comparison. When finding saddle points, use a test system where the energy surface can be visualized and the search path illustrated. Also, when it comes to implementation of a previously developed approach, do a calculation on a system that has been studied with the method previously and where the performance has been documented. Here, the web site OptBench.org is of great help. There, various benchmark problems have

been documented and the performance of various methods reported by experts. There are too many articles in the literature where performance of a method is reported but the implementation is not optimal. It is important to compare performance reported by those who developed or are knowledgeable about the method.

The challenge in atomic scale simulations is typically the large number of degrees of freedom. While a system with only a few degrees of freedom can be valuable to test and illustrate a method, performance should not be measured with application to such systems. A method that works well for a system with only a few degrees of freedom may not work well for a realistic system with many degrees of freedom. The opposite can also be the case. It is important to document performance on challenging systems for which the method has been developed.

When documenting performance, it does not make sense to report CPU time. Computers change rapidly and such information is quickly obsolete. Identify the most computationally intensive operation and report the number of such operations needed to reach the desired results from the simulation. In most simulations of atomic scale systems, the evaluation of the energy and atomic forces is the most computationally intensive operation. A natural measure of computational effort in calculations of transition rate, identification of a reaction mechanism or a saddle point search is the number of times the energy and force needs to be evaluated.

It is also important to keep focus on the ultimate goal of the calculation, not intermediate steps. The question whether an approach is useful and how large the computational effort is should be answered by evaluating the quantity of interest.

Performance in terms of computational effort is of course not the only criterion for evaluating the appropriateness of a method. A method that is not reliable in that the answer obtained cannot be trusted is not useful even if it is fast. The results obtained using a fast method should be compared with results obtained using a slower but safer method on a range of problems similar to the application of interest.

GH: The points raised, including the recommendations to compare new methods to existing methods through benchmarks and the strategy of developing methods on model system with known results and then demonstrating how well they work in complex high dimensional systems are spot-on. Adopting these recommendations as best practices would benefit the community of method developers and the people who aim to apply the methods. An additional recommendation, which is gaining traction in the community, is that computational methods that are developed in the public domain should be made available in the form of open source code. There are many details associated with computational methods which are not easily described in publications. To make our methods transparent and our calculations reproducible, other developers and users should be able to see the algorithms at the level of the source code and reproduce published results and benchmarks directly from the code. The adoption of an open-source policy for computational material science is, in my opinion, encouraging collaboration in the field and facilitating the development of computational frameworks that are larger than the scope of a single research group, such as the Materials Project.

Finally, I think that there is an opportunity in the field to consider how computational methods compare with respect to more than one objective. Taking the example of modeling molecular dynamics over long timescales, for example, we have efficient methods based upon harmonic transition state theory (e.g., temperature accelerated dynamics and off-lattice kinetic Monte Carlo), which have the inherent limitations associated with that approximation, and other methods for which the harmonic approximation is relaxed (e.g., parallel replica dynamics and hyperdynamics), that may have additional computational costs. What is not typically considered is the pareto-optimal set of methods which can deliver the highest accuracy for the minimum computational cost. In other words, the community of developers can establish a set of tools which can most efficiently accelerate dynamics for a specified level of accuracy. There is similar multiobjective optimization problem between the accuracy of the energy and force evaluations (e.g., empirical potentials vs. density functional theory) and the degree to which the energy landscape can be sampled. Research groups that focus on the accuracy of electronic structure calculations can neglect the potential importance of exploring the energy landscape. On the other hand, a focus on highly accurate sampling will typically put little emphasis on the accuracy of the potential landscape. A set of efficient methods for modeling dynamics or sampling potential surfaces targeting a wide range of available sample sizes would facilitate the use of methods such as accelerated dynamics for the large community of scientists modeling systems of interest with density functional theory.

AFV: Graeme and Hannes have covered this well; I will add just a few general points. As with any careful computational work (or any careful science for that matter), one should always be on the lookout for indications that something is not working correctly or not making sense. This is especially important for simulation methods that are capable of giving results for timescales that cannot be reached in any other way. For example, perhaps there is a nonlinearity that only causes significant inaccuracy at very large boost factors, which means it might not show up until the simulated timescales are beyond what can be checked with MD.

Although I think most readers of this chapter will already understand this, there is a general principle I feel is important for developing any method of this type – a method that is tied to an interatomic potential and that attempts to improve the efficiency of the simulation of a physical, material, or chemical property. When testing the method, the accuracy should be gauged by how well the method reproduces accurate simulation results for that same interatomic potential, *not* an experimental result. For long-time dynamics methods, the correct reference is direct molecular dynamics, and the benchmark systems must be chosen with some care, as MD may not be capable of directly reaching the long timescales (as Hannes pointed out).

Finally, on this issue of common codes, while I agree that standard software packages make development faster and easier than ever, one should absolutely not be afraid of developing one's own code from scratch. This can take longer, but sometimes it opens possibilities for creativity in the development that would be steered in a different direction, or inhibited, by using existing packages. Moreover,

the developer typically gains a deeper understanding of the methodology by proceeding in this way. And standard codes are not totally bug- or mistake-free, so the developer may uncover such problems by having a redundant code with which to compare, thereby doing the community a favor.

NM: Previous advice is excellent and I agree with all those. Developing methods is a risky business. It is generally impossible to tell how well it will perform on realistic systems before the method is in place. For a method to be useful, it has to be beyond what is available, either by doing faster or allowing access to new regimes of physics. This is why, I would also suggest that you make sure that the method that you are developing can go beyond what other approaches can do by applying it, as soon as possible, on nontrivial problems. So: yes, do check on simple systems that can allow you to compare results with other approaches, but make sure that also apply your algorithm to something nontrivial that demonstrate the interest of your method; too many methods are demonstrated on simple well-understood systems but have failed to produce any new significant physics.

It is also worth remembering that few computational physicists today write extensive codes. Most will use standard packages and write analysis or extension bits. If your algorithm has any degree of complexity, it is therefore essential that you write a code that is portable and useable by others, if you want your approach to gain exposure and have the impact you expect. This means that your code should be easy to read, modular, and adaptable. Be aware of this requirement from the day you start planning your software; this will decrease the time you will have to spend getting your code ready for distribution.

TL: Let me make two points along those lines. First, I would like to mention here that it is actually very difficult for nonspecialists to have access to test problems which are simple enough to have reference values, but not considered overly simple by the applied community. It would be very useful for the development of new approaches to agree on a sequence of test problems with graded difficulty, where the problem would be only a timescale problem (and not a modeling problem related to the force field for example). Second, when very long timescales are reached, weird results may be observed because of two reasons: (i) the algorithm gives incorrect estimates or (ii) the model is incorrect. Concerning the second item, notice indeed that many force fields have been parameterized and checked using only short-timescales simulations: when looking at very long trajectories, one is thus visiting unexplored territories. In such situations, it is very important to have rigorous ways to assess the quality of the numerical results in order to distinguish between the two sources of errors. This shows the importance of deriving certified error bounds for such algorithms (which is indeed sometimes a challenge!).

# References

Steiner MM, Genilloud PA, Wilkins JW (1998) Simple bias potential for boosting molecular dynamics with the hyperdynamics scheme. Phys Rev B 57:10236

# Part VI
# Modeling Tools for Magnetism, Magnetic Materials, and Spintronics

# Modeling Tools for Magnetism, Magnetic Materials, and Spintronics: Overview

# 39

Stefano Sanvito

## Contents

**Abstract**

Spanning over a wide range of length and time scales, magnetism is one of the most fascinating states of matter and one that poses many challenges to theory. Here we briefly review how different state-of-the-art approaches to the modeling of the magnetic order interface with each other and provide a complete toolset to explore this fascinating area of condensed matter theory. Methods include accurate quantum-mechanical approaches, such as density functional theory both in its static and dynamical version, atomistic spin dynamics tools, and continuum models to be solved in a finite difference framework.

## 1    Modeling Magnetism

Magnetism is one of the most fascinating macroscopic orders of matter. It is deeply rooted in quantum mechanics and spans over many different scales in time, space, and energy (see Fig. 1). At its core magnetism is based on the $m$-$J$-$\kappa$ paradigm. The Hund's coupling is responsible for the formation of the magnetic moment, $m$, often localized close to the atomic nuclei. This is active in open shell ions; hence

S. Sanvito (✉)
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Center for Materials Genomics, Duke University, Durham, NC, USA
e-mail: sanvitos@tcd.ie
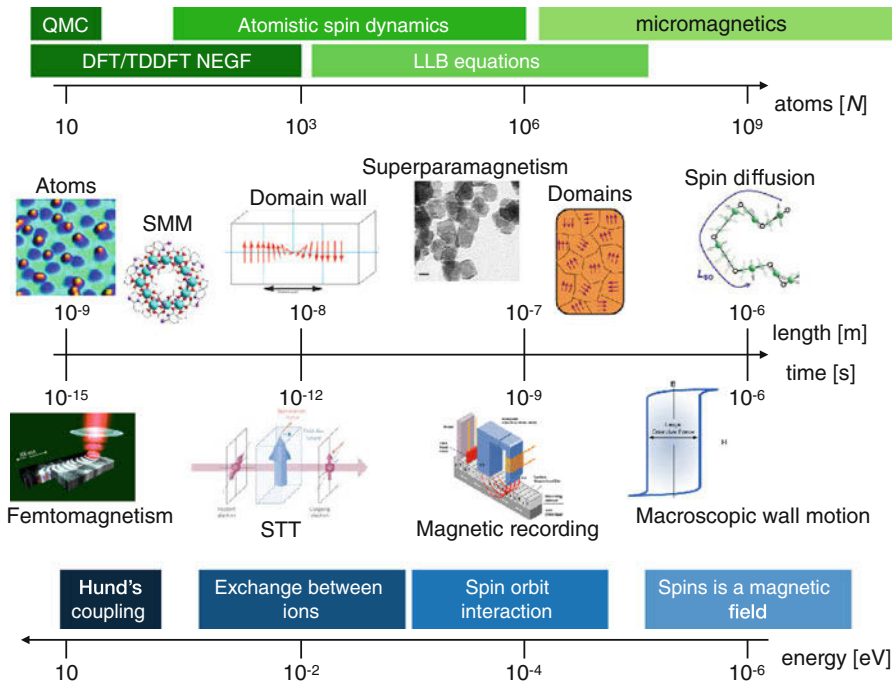
**Fig. 1** Length, time and energy scales of magnetism. At the atomic level, the high-energy Hund's coupling is responsible for the formation of the local moments, which are then coupled to each other through exchange interaction. Finally, magnetic anisotropy aligns the moment along particular directions in space. The computational tools needed to explore and explain magnetism range from quantum mechanics-rooted methods, such as density functional theory (DFT) and Quantum Monte Carlo (QMC), to atomistic spin dynamics, and to effective theories allowing large-scale coarse graining, such as the Landau-Lifshitz-Bloch (LLB) and the micromagnetic approach

$m$ forms in $2p$ radicals, among the elements in the $3d$ period, in rare earths and some $4d$ ions. Then the exchange interaction, $J$, makes the moments interacting with each other. At the microscopic level, a large variety of mechanisms is active in generating $J$ at both sides of the metal/insulator boundary. These range from direct exchange to ion-mediated mechanisms and to those where the magnetic interaction is mediated by free electrons. Finally, the magnetic anisotropy, $\kappa$, is responsible for the interplay between spatial and spin degrees of freedom and ties the local moment to specific directions in space. The anisotropy, at the atomic level, is determined by the spin-orbit coupling.

The interplay and competition among the different interactions participating to the $m$-$J$-$\kappa$ paradigm guide the formation of a multitude of magnetic orders. Locally exchange and anisotropy contribute to establish an alignment among the local moments, thus that often a microscopic order emerges. In simple cases, this can have the same periodicity of the crystal unit cell, but most frequently the crystal and magnetic cells are different. Then, classical magnetostatics imposes the formation

of macroscopically ordered regions (domains), separated by others where the order changes (domain walls). Symmetry breaking, for example through the presence of a surface, introduces additional complexity. In general the exchange and anisotropy get altered and, as a result, so does the macroscopic order.

Excitations of different energy can tip the various interactions and probe magnetism over many time scales. Intense, femtosecond-long, laser pulses interact directly with the electronic degrees of freedom (femtomagnetism). At this energy and time scale, the Hund's coupling, the exchange interaction, and the magneto-crystalline anisotropy are all functions of time, and the short-time dynamics is determined by their evolution. As time passes, the most energetic of the interactions, the Hund's coupling, returns $m$ to its ground state value, meaning that the local moment recovers first. The dynamics is then determined by electronic currents interacting with local spins via exchange interaction. The same dynamics can also be engineered by passing spin-polarized electrical currents through magnets of various order. This is the realm of spin-transfer-torque devices. Now the exchange and the anisotropy are perturbed, and the magnetic order can be manipulated, both locally and globally. The energy scale of this phenomena is lower than that needed to alter the local moment, and the typical times vary between picoseconds and nanoseconds. In the slower limit of this range operates magnetic data storage technologies. Finally, when the dynamics is governed by macroscopic domain walls motion, magnetism slows down further, the exchange interaction becomes constant, and the time scale may approach the macroscopic range.

It is then clear that magnetism is both an atomic and a macroscopic phenomenon, and as such, it necessitates computational tools spanning a broad range of time and length scales. This chapter wishes to provide a map of the theoretical methodologies used to date to describe and understand the magnetic interaction.

At the atomic level, density functional theory (DFT) is usually the theory of choice, since it can describe all the interactions on the same footing and does not require any parameters from experiments. DFT for magnetism will be reviewed in ▶ Chap. 42, "Density Functional Theory for Magnetism and Magnetic Anisotropy" by Bihlmayer, while Friedrich, Müller, and Bluegel in ▶ Chap. 43, "Spin Excitations in Solid from Many-Body Perturbation Theory" will present how it can be applied to the description of magnetic excitations. Beyond-DFT methods usually require computational resources significantly larger than DFT itself. However, in some cases, they provide a more accurate description of both the moment formation and the exchange. Recent progress in Monte Carlo technique for $d$ and $f$ ions will be reviewed by Wagner in ▶ Chap. 46, "Quantum Monte Carlo for Electronic Systems Containing $d$ and $f$ Electrons". Finally, the DFT description of magnetism can be extended to time-dependent phenomena, as discussed by Elliott and co-workers in ▶ Chap. 40, "Time-Dependent Density Functional Theory for Spin Dynamics", and to open systems in the presence of a steady-state current, as described by Rungger and collaborators in ▶ Chap. 44, "Non-equilibrium Green's Function Methods for Spin Transport and Dynamics".

The next level of description is obtained by neglecting an explicit description of the electronic degrees of freedom. Namely, one can represent a magnet by

using local spins of constant magnitude associated to each atom. This provides an atomistic view of spin phenomena, at a course-graining scale analogous to that used in classical molecular dynamics for the ionic motion. These methods, reviewed by Ma and Dudarev in ▶ Chap. 47, "Atomistic Spin-Lattice Dynamics", can describe magnetic phase transitions for systems including up to several millions atoms and require parameters extracted either from experiments or from DFT calculations. Intriguingly, the interplay between the spin and the vibrational degrees of freedom can be described on the same footing. One can then integrate the dynamics over larger regions of space and move to a continuous description. This is the domain of micromagnetic methods, which allow to reach time and length scales approaching the macroscopic limit. The foundation of micromagnetic theory and and its connection to spintronics are discussed by Abert in ▶ Chap. 45, "Spintronics in Micromagnetics". In its most conventional form, only the transverse component of the magnetization can vary, while its longitudinal part is constant. However, a formalism based on the Landau-Lifshitz-Bloch equations allows to go beyond this approximation. Such range of methods is presented by Chubykalo-Fesenko and Nieves in ▶ Chap. 41, "Landau-Lifshitz-Bloch Approach for Magnetization Dynamics Close to Phase Transition".

## 2 Conclusion

We have briefly reviewed the state of the art in the modeling of the magnetic order both in the static and time-dependent domains. This develops over a range of computational schemes capable of bridging many length and time scales. In particular we have discussed the approximations that allow one to connect the various levels of theory, where a progressively smaller number of degrees of freedom are treated explicitly. Such a body of work gives us a complete toolset to tackle this fascinating area of materials science.

## References

Steiner MM, Genilloud PA, Wilkins JW (1998) Simple bias potential for boosting molecular dynamics with the hyperdynamics scheme. Phys Rev B 57:10236

## Peter Elliott, Maria Stamenova, Jacopo Simoni, Sangeeta Sharma, Stefano Sanvito, and Eberhard K. U. Gross

## Contents

P. Elliott (✉) · S. Sharma · E. K. U. Gross
Max Planck Institute of Microstructure Physics, Halle, Germany
e-mail: peter.elliott@mpi-halle.mpg.de; sangeeta.sharma@mpi-halle.mpg.de; hardy@mpi-halle.mpg.de

J. Simoni
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: jsimoni@lanl.gov

M. Stamenova
School of Physics, AMBER and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: stamenom@tcd.ie

S. Sanvito
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland

Center for Materials Genomics, Duke University, Durham, NC, USA
e-mail: sanvitos@tcd.ie

**Abstract**

With the development of ultrashort sub-picosecond laser pulses, the last two decades have witnessed the emergence of a new field of magnetism, namely, femtomagnetism. This consists of controlling the magnetic interactions by using purely optical stimuli at sub-picosecond timescales, where both the exchange interaction and the magnetic anisotropy cannot be considered constant. The modeling of such phenomena is at present populated by semiempirical theories, which heavily rely on assumptions about the dominant interactions responsible for the dynamics and the system intrinsic properties (e.g., the conductivity). However, in the last few years, there have been a few attempts to look at the problem from a purely ab initio point of view, namely, by using time-dependent density functional theory. Here we will review the progress in this field and show how a theory not biased by assumptions can shed light into the fundamental aspects of the laser-induced magnetization dynamics. In particular we will discuss the ultrafast demagnetization of transition metals both in their cluster and bulk form and the possibility of spin transfer between sublattices in compounds containing magnetic ions. The chapter is also complemented by a short review of time-dependent spin density functional theory in the context of spin dynamics.

# 1    Introduction

In 1996 it was first demonstrated that an intense laser pulse could generate ultrafast light-induced demagnetization (in Ni), with typical demagnetization times faster than a few picoseconds (Beaurepaire et al. 1996). Recently, these spin-manipulation times have been reduced to a few femtoseconds, owing to the great advances made in the production of short-time laser pulses. As the control of spin moments by light may strongly impact several technological applications such as magnetic storage, spintronics, all-optical switching, heat-assisted magnetic recording, just to name a few, the field of femtomagnetism has recently become intensively active. To date several experimental methods are available to study magnetism at high speed. These include THz emission (Kampfrath et al. 2011; Walowski and Münzenberg 2016), time-resolved X-ray circular dichroism (Stamm et al. 2007; Boeglin et al. 2010), and high harmonic generation (La-O-Vorakiat et al. 2009).

Femtomagnetism has two ultimate goals: (i) to understand the physics of light-induced spin dynamics and, on this basis, (ii) to manipulate the spin in a fully controlled manner using designed materials and/or tailored laser pulses. However, a complete understanding of the underlying physics of light-induced spin dynamics is hard to reach. In fact, since there is not strong direct coupling between light and spins, laser-induced spin dynamics is an indirect phenomenon, driven by spin-charge and spin-lattice interactions. A variety of distinct and simultaneous physical mechanisms thus underpins the observed magnetization dynamics. Some of these, such as interatomic spin transfer or spin-flips, begin early after the laser pulse, while other processes such as spin canting, spin currents (spin diffusion), or spin-lattice-induced spin-flips make significant contributions only at a later time (Fig. 1).
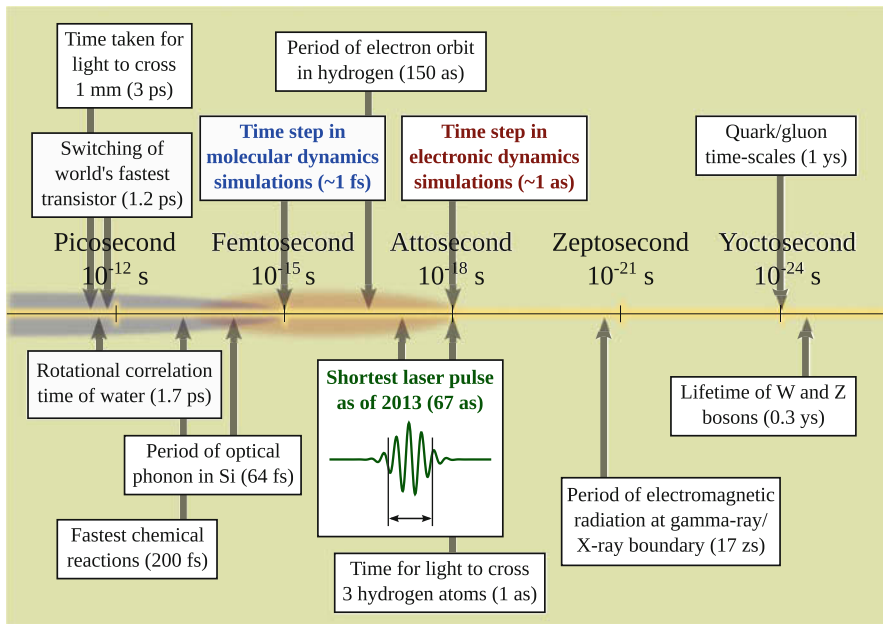
**Fig. 1** Comparison of timescales for sub-nanosecond physical phenomena, where we have also included the characteristic times for time-dependent molecular and electronic dynamics

Disentangling these processes and identifying the relevant timescales are among the first outstanding challenges of this field.

## 2    Models of Ultrafast Spin Dynamics

There have been several distinct attempts at explaining optically induced spin dynamics. At the thermodynamical level, three-temperature models (3TMs), which assume the ultrafast demagnetization to be a thermal process, are widely used (Beaurepaire et al. 1996). For times earlier than picoseconds, the magnetic system cannot be considered in thermal equilibrium with electrons and phonons. The laser-induced initial hot electron distribution initiates spin dynamics, which is driven by a spin temperature different from the electronic and lattice ones. Thus the electronic and lattice degrees of freedom act as energy reservoirs for the spin system. From a practical point of view, one solves three coupled differential equations representing the electronic, spin, and lattice dynamics. The demagnetization then crucially depends upon the coupling strengths between the three subsystems, coupling strengths that remain free parameters. Fitting is performed against experimental data usually with good success (Vodungbo et al. 2012; Ostler et al. 2012).

At the dynamical level, there is a vast range of approaches depending on the material type and conditions. These include lattice models using the Landau-Lifshitz-Gilbert equation with effective magnetic fields and damping parameters (Kazantseva et al. 2008) and direct spin-photon interaction models derived from

a lower-order nonrelativistic expansion of the Dirac Hamiltonian (Hinschberger and Hervieux 2012; Jean-Yves Bigot and Beaurepaire 2009). This type of description usually neglects many-body electron-electron interactions by assuming that during the application of the laser pulse, the demagnetization process occurs through the interplay between the spin-laser and spin-orbit interactions.

When electron transport takes place, the Elliott-Yafet mechanism describes scattering-induced spin-flips (Koopmans et al. 2010) modulated by phonons, which may lead to a global momentum loss. The speed of such demagnetization process is, however, bound to the slow lattice dynamics and hence cannot be used to explain early-femtosecond ($\sim$10–20 fs) demagnetization. The super-diffusive spin-transport mechanism (Battiato et al. 2010) instead completely neglects spin-orbit coupling (SOC) and relies on spins diffusing away and thus causing a local reduction in the material moment. These assumptions are controversial, and the literature is populated by experimental results confirming them (Vodungbo et al. 2012) and others contesting their validity (Schellekens et al. 2013).

## 3    Why Time-Dependent Spin Density Functional Theory?

Given the wide range of theoretical approaches and the uncertainty over their assumptions and range of validity, it is desirable to look into an ab initio method to describe ultrafast spin dynamics. Time-dependent density functional theory (TDDFT) is an in-principle exact approach for calculating electron, and thus spin, dynamics induced by external fields. It builds on ground-state (GS) density functional theory and has been successfully applied in the linear-response regime. We will first review the basic foundations of TDDFT before detailing several recent applications of TDDFT to spin dynamics. For simplicity we will introduce TDDFT for the spin-unpolarized case and use atomic units throughout unless otherwise specified.

### 3.1    Basic TDDFT Concepts

We begin with the nonrelativistic time-dependent Schrödinger equation:

$$i \frac{\partial}{\partial t} \Psi(\mathbf{r}_1, \mathbf{r}_2 \ldots \mathbf{r}_N, t) = \hat{H}(t) \Psi(\mathbf{r}_1, \mathbf{r}_2 \ldots \mathbf{r}_N, t), \qquad \Psi(t_0) \text{ given,} \qquad (1)$$

which governs the dynamics of $N$ electrons, starting from a given initial state, $\Psi(t_0)$, and under the influence of a time-dependent Hamiltonian, $\hat{H}(t)$. In the spatial representation, the Hamiltonian is written as

$$\hat{H}(t) = \sum_{i=1}^{N} \left[ -\frac{1}{2} \nabla_i^2 + v_{\text{ext}}(\mathbf{r}_i, t) \right] + \frac{1}{2} \sum_{i \neq j}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \qquad (2)$$

where the first term is the kinetic energy, the last is the electron-electron interaction, and $v_{ext}(\mathbf{r}, t)$ is the time-dependent external potential, which contains both the electron-nuclear interaction and any external electric fields (e.g., a laser pulse). The time-dependent one-electron density is defined as

$$n(\mathbf{r}, t) = N \int d^3 r_2 \ldots \int d^3 r_N |\Psi(\mathbf{r}, \mathbf{r}_2, \ldots \mathbf{r}_N, t)|^2 . \tag{3}$$

TDDFT is based on the Runge-Gross (RG) theorem (Runge and Gross 1984), which proves that the knowledge of $n(\mathbf{r}, t)$, rather than the full wavefunction, is sufficient to determine all the observables of the system. The RG theorem states that, if two time-dependent Taylor-expandable (around the initial time $t_0$) external potentials, $v_{ext}(\mathbf{r}, t)$ and $v'_{ext}(\mathbf{r}, t)$, differ by more than a purely time-dependent function, then the associated time-dependent densities, $n(\mathbf{r}, t)$ and $n'(\mathbf{r}, t)$, evolving from a common initial state, must be different. In other words, there is a one-to-one mapping between the density and the external potential and hence all the observables.

The proof of the RG theorem is split into two parts. First we consider the difference $\Delta\mathbf{j}(\mathbf{r}, t) = \mathbf{j}(\mathbf{r}, t) - \mathbf{j}'(\mathbf{r}, t)$ between two current densities evolving in the potentials $v_{ext}(\mathbf{r}, t)$ and $v'_{ext}(\mathbf{r}, t)$, respectively. The time evolution of this difference is

$$\left.\frac{\partial \Delta\mathbf{j}(\mathbf{r}, t)}{\partial t}\right|_{t=0} = -n_0(\mathbf{r})\nabla\Delta v_{ext}(\mathbf{r}, 0) , \tag{4}$$

where $\Delta v_{ext}(\mathbf{r}, 0) = v_{ext}(\mathbf{r}, 0) - v'_{ext}(\mathbf{r}, 0)$ and $n_0(\mathbf{r})$ is the initial density (which is the same for both systems due to the common initial wavefunction). Equation (4) was derived by inserting the paramagnetic current density operator:

$$\hat{\mathbf{j}}_p(\mathbf{r}) = \frac{1}{2i} \sum_{j=1}^{N} \left[\nabla_j \delta(\mathbf{r} - \mathbf{r}_j) + \delta(\mathbf{r} - \mathbf{r}_j)\nabla_j\right] , \tag{5}$$

into the Heisenberg equation of motion. At this point, we make use of the restriction to the class of external potential, which are Taylor expandable around the initial time, $t_0$, namely:

$$v_{ext}(\mathbf{r}, t) = \sum_{j=0}^{\infty} \frac{1}{j!} v_j(\mathbf{r})(t - t_0)^j , \tag{6}$$

where $v_j(\mathbf{r})$ are the Taylor coefficients of this expansion. If $v_{ext}(\mathbf{r}, t)$ and $v'_{ext}(\mathbf{r}, t)$ differ by more than a time-dependent constant, then there must be some order, $k$, at which the Taylor expansion coefficient of $\Delta v_{ext}(\mathbf{r}, t)$ is not a constant

$$u_k(\mathbf{r}) - u'_k(\mathbf{r}) = \left. \frac{\partial^k}{\partial t^k} \left[ v_{\text{ext}}(\mathbf{r}, t) - v'_{\text{ext}}(\mathbf{r}, t) \right] \right|_{t=t_0} \neq \text{constant} . \tag{7}$$

Thus, if we take $k$ additional time derivatives of Eq. (4), we can see that the time evolution of the two current densities must be different. This establishes the one-to-one correspondence between external potential and the current density. The second step of the RG proof utilizes the continuity equation:

$$\frac{\partial n(\mathbf{r}, t)}{\partial t} = -\mathbf{\nabla} \cdot \mathbf{j}(\mathbf{r}, t) , \tag{8}$$

along with Eq. (4) to relate the charge density to the external potential and prove that the mapping between them is unique. Thus we can write $v_{\text{ext}}[n, \Psi_0](\mathbf{r}, t)$, i.e., the external potential is a functional of the density and the initial wavefunction.

We can also apply the RG theorem to a noninteracting system, referred to as the Kohn-Sham (KS) system. There is then a potential, $v_{\text{S}}[n, \Phi_0](\mathbf{r}, t)$, such that the evolution of noninteracting fermions in this potential reproduces the exact time-dependent density. This requires various assumptions about the $N$- and $v$-representability of the density and potential, a discussion that is beyond the scope of this review, but can be explored in van Leeuwen (1999), Li and Ullrich (2008), and Ruggenthaler and van Leeuwen (2011). The time-dependent Kohn-Sham equations are

$$i \frac{\partial \phi_j(\mathbf{r}, t)}{\partial t} = \left[ -\frac{\nabla^2}{2} + v_0(\mathbf{r}, t) + v_{\text{H}}[n](\mathbf{r}, t) + v_{\text{xc}}[n](\mathbf{r}, t) \right] \phi_j(\mathbf{r}, t) , \tag{9}$$

where $\phi_j(\mathbf{r}, t)$ denotes the single-particle Kohn-Sham orbitals. The density is given by:

$$n(\mathbf{r}, t) = \sum_{j=1}^{N/2} \left| \phi_j(\mathbf{r}, t) \right|^2 , \tag{10}$$

and the Hartree potential is

$$v_{\text{H}}[n](\mathbf{r}, t) = \int d^3 r' \frac{n(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} . \tag{11}$$

The exchange-correlation (XC) potential, $v_{\text{XC}}$, is defined as

$$v_{\text{xc}}[n, \Psi_0, \Phi_0](\mathbf{r}, t) \equiv v_{\text{s}}[n, \Phi_0](\mathbf{r}, t) - v_{\text{ext}}[n, \Psi_0](\mathbf{r}, t) - v_{\text{H}}[n](\mathbf{r}, t) \tag{12}$$

and has a functional dependence on the density of the system at the current and at all previous times, as well as the initial interacting and noninteracting wavefunctions. These initial states need not be the GS or an eigenstate of the initial potential,

$v_{\text{ext}}(\mathbf{r}, t_0)$. However, if we start from the GS, then all initial-state dependence is subsumed into the initial density due to the Hohenberg-Kohn theorem (Hohenberg and Kohn 1964). In practice, the XC potential must be approximated, with the most commonly used being the adiabatic approximation. This simply inserts the density at each time into a ground-state DFT XC functional.

## 3.2   TDSDFT for Ultrafast Spin Dynamics

TDDFT can be extended to include the electron spin degree of freedom and its interaction with an external magnetic field, $\mathbf{B}_{\text{ext}}(\mathbf{r}, t)$, via the Zeeman interaction. Hence, in addition to the exact density $n(\mathbf{r}, t)$, TD spin DFT (TDSDFT) will also reproduce the exact magnetization (or spin) density, $\mathbf{m}(\mathbf{r}, t) = \text{Tr}\left[\hat{\rho}(\mathbf{r}, t)\boldsymbol{\sigma}\right]$, where $\hat{\rho}$ is the single-particle density matrix, i.e., $n(\mathbf{r}, t) = \text{Tr}\left[\hat{\rho}(\mathbf{r}, t)\right]$, and $\boldsymbol{\sigma}$ is the vector of Pauli matrices. The TD-KS equations become (from Eq. (9)):

$$
i\frac{\partial \phi_j(\mathbf{r}, t)}{\partial t} = \left[\frac{1}{2}\left(-i\boldsymbol{\nabla} + \frac{1}{c}\mathbf{A}_{\text{ext}}(t)\right)^2 + v_{\text{S}}(\mathbf{r}, t) + \frac{1}{2c}\boldsymbol{\sigma}\cdot\mathbf{B}_s(\mathbf{r}, t)\right.
$$
$$
\left. + \frac{1}{4c^2}\boldsymbol{\sigma}\cdot(\boldsymbol{\nabla}v_s(\mathbf{r}, t) \times -i\boldsymbol{\nabla})\right]\phi_j(\mathbf{r}, t) , \tag{13}
$$

where $\mathbf{B}_{\text{S}}(\mathbf{r}, t) = \mathbf{B}_{\text{ext}}(\mathbf{r}, t) + \mathbf{B}_{\text{XC}}(\mathbf{r}, t)$ is the KS effective magnetic field and $\mathbf{B}_{\text{XC}}$ is the XC magnetic field, which may be approximated for non-collinear systems according to the method of Kübler (Kubler et al. 1988). In Eq. (13), the external laser field is written as a purely time-dependent vector potential, $\mathbf{A}_{\text{ext}}(t)$. If gauge transformed to the length gauge, it recovers the dipole approximation to the electric field of a laser. This form is necessary for periodic systems so that the Hamiltonian retains lattice periodicity. The last term of Eq. (13) is the SOC. If we assume $v_{\text{S}}(\mathbf{r}, t)$ to be spherically symmetric around each atom, then this term reduces locally to the more familiar form, $\xi(r)\,\hat{\mathbf{L}}_e \cdot \hat{\mathbf{S}}$, where $\hat{\mathbf{L}}_e$ is the electronic angular momentum operator, $\hat{\mathbf{S}} = 1/2\hat{\boldsymbol{\sigma}}$ is the spin operator, and $\xi(r) = \partial_r v_{\text{S}}(r)/2c^2 r$. SOC typically provides small contributions to the GS energy, but, as we will see, it has dramatic effect on the dynamics.

### 3.2.1   Spin Continuity Equation
A spin continuity equation for the local magnetization density can be derived using the Heisenberg equation of motion (EOM) along with the Hamiltonian in Eq. (13):

$$
\frac{d}{dt}\mathbf{m}(\mathbf{r}, t) = -\boldsymbol{\nabla}\cdot\overset{\leftrightarrow}{\mathbf{J}}_{\text{S}}(\mathbf{r}, t) + \frac{1}{2c}\mathbf{m}(\mathbf{r}, t) \times \mathbf{B}_s(\mathbf{r}, t) + \frac{1}{4c^2}\langle(\boldsymbol{\nabla}v_s(\mathbf{r}) \times -i\boldsymbol{\nabla}) \times \hat{\boldsymbol{\sigma}}\rangle \tag{14}
$$

where the first term on the right-hand side represents the spin current divergence of the noninteracting Kohn-Sham system. Here we defined the spin current operator to be $\overset{\leftrightarrow}{\hat{\mathbf{J}}}_{\text{S}}(\mathbf{r}, t) = \hat{\boldsymbol{\sigma}} \otimes \hat{\mathbf{j}}$ or, equivalently, $\hat{\mathbf{j}}_\nu = \hat{\sigma}_\nu\hat{\mathbf{j}}$, where $\nu = \{x, y, z\}$ and

$\hat{\mathbf{j}}(\mathbf{r}, t) = \hat{\mathbf{j}}_p(\mathbf{r}) + \hat{n}(\mathbf{r}) A_{\text{ext}}(t)/c$ is the usual current density operator (which includes a diamagnetic term in order to be gauge invariant). The middle term is a precessional term due to the effective KS magnetic field, and the last term is the magnetization torque due to SOC.

We may integrate Eq. (14) to find the equation of motion for the global spin moment, $\mathbf{M}(t) = \int d^3r \; \mathbf{m}(\mathbf{r}, t)$. After some algebra and in the absence of an external magnetic field, one can write:

$$\frac{\partial}{\partial t}\mathbf{M}(t) = \frac{1}{2c^2} \int d^3r \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\nabla} v_s(\mathbf{r}, t) \times \mathbf{j}_x(\mathbf{r}, t) \\ \boldsymbol{\nabla} v_s(\mathbf{r}, t) \times \mathbf{j}_y(\mathbf{r}, t) \\ \boldsymbol{\nabla} v_s(\mathbf{r}, t) \times \mathbf{j}_z(\mathbf{r}, t) \end{bmatrix} , \tag{15}$$

where we used Gauss' theorem and the TDSDFT zero-torque theorem (Capelle et al. 2001). Equation (15) demonstrates that only SOC can change the global spin moment.

## 4 Ultrafast Demagnetization of Small Magnetic Clusters

We will now review some of the results obtained with TDSDFT towards the understanding of ultrafast spin dynamics by starting from transition-metal clusters and then moving to bulk magnets. TDSDFT simulations have shown that, similarly to ferromagnetic films, a number of small iron clusters (i.e., $Fe_2$, $Fe_4$ and $Fe_6$) demagnetize rapidly when excited by a strong electric field pulse a few femtosecond long (Stamenova et al. 2016). The real-time evolution of the system described by TDSDFT simulations allows one to gain unparalleled microscopical insights into the physical processes, driving the ultrafast dissipation of spin angular momentum. It is understood that, in the absence of direct light-to-spin coupling, the demagnetization onset is triggered by the electronic charge response to the electric field component of the laser pulse. A more transparent microscopic interpretation of the observed first-principles spin dynamics can be sought in the spin continuity equation, which in turn leads to a set of coupled magnetohydrodynamical equations (Simoni et al. 2017). This will allow us to understand the demagnetization process through the following mechanism: the charge and spin currents generated by the electric pulse in the magnetic media give rise to an effective magnetic field that, in combination with the SOC, facilitates spin-flip and a global spin decay. We will elaborate on this picture bellow.

### 4.1 TDSDFT Simulations

The TDSDFT platform of choice is the OCTOPUS code (Castro et al. 2006), which uses a real-space representation of the wavefunctions and is particularly suited for an accurate treatment of small atomistic systems and shaped time-dependent
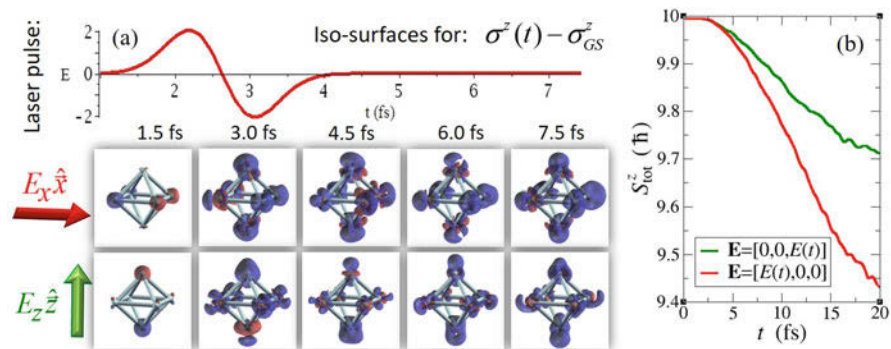
**Fig. 2** (**a**) A typical pulse used in the TDSDFT simulation and isosurface snapshots of the spin density variation with respect to the GS at given times after the initiation of the pulse for two different directions of the electric field relative to the cluster geometry. Red and blue isosurfaces correspond, respectively, to positive and negative variation of the same magnitude with respect to the GS spin density. (**b**) The evolution of the net spin of the cluster for the two directions of the electric field

excitations. SOC is included at the level of pseudopotential, $v_{PP}(r)$, i.e., it is part of the KS potential, $v_0(r)$, of Eq. (9):

$$v_{PP}(r) = \sum_I \sum_l \sum_{m=-l}^{l} \left( \bar{V}_l^{ion}(r) + \frac{1}{4} V_l^{SO}(r) + \alpha V_l^{SO}(r) \hat{\mathbf{L}}_I \cdot \hat{\mathbf{S}} \right) |l, m\rangle \langle l, m| \, , \tag{16}$$

where $\hat{\mathbf{S}}$ is the spin operator, $\hat{\mathbf{L}}_I$ is the angular momentum operator associated to the atomic center "$I$," while the scalar part of the pseudopotential, $\bar{V}_l^{ion}(r)$, includes the effect of the mass shift and the Darwin term. Here $V_l^{SO}(r)$ defines the range of the SOC term, and $\alpha$ is a scaling parameter used to enhance/reduce the contribution of the SOC from the regular value $\alpha = 1$. In the following the adiabatic local spin density approximation (ALSDA) to the XC functional has been used with the Perdew and Wang parameterization.

Any TDSDFT simulation starts with a GS SDFT calculation. We then consider a strong bipolar pulse of an homogeneous electric field as illustrated in Fig. 2a and applied as per Eq. (13) in length gauge. In the case of $Fe_6$, which has the shape of an (almost) square-based bipyramid (Gutsev and Beuschlicher 2003), the direction of the electric field with respect to the cluster affects the net demagnetization process. The snapshots from such TDSDFT simulations, presented in Fig. 2, highlight the local variation of the spin density of $Fe_6$ for two pulses differing only by the E-field direction. It is clear that at the pulse onset, there is charge transfer, accompanied by spin transfer (as the cluster is spin-polarized), in the direction of the electric field. Soon after that a spin decay for the entire cluster is established (the predominant color in the plots is blue). For both field directions, this results in demagnetization, which continues long after the pulse has died out (Fig. 2b). For $Fe_6$ the effect is
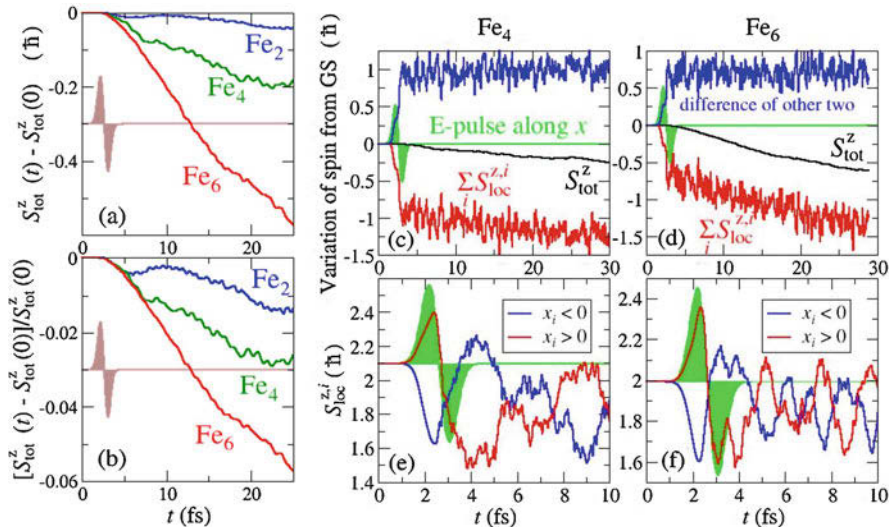
**Fig. 3** Comparison of the absolute (**a**) and relative (i.e., per unpaired electron in the cluster) (**b**) demagnetization in three Fe clusters: $Fe_2$, $Fe_4$, and $Fe_6$. Panels (**c**) and (**d**) show the evolution of the total spin of the cluster (black curve) compared to the *local spin*, integrated over spheres of radius $r_{sph} = 0.5$ Å (red curves) centered at the atoms (see text) and the difference of the latter. Panels (**e**) and (**f**) display a comparison of average local spins for the two pairs of atoms in the plane along the direction of the pulse ($x$-axis) for $Fe_4$ and $Fe_6$, respectively

very anisotropic with the cluster demagnetizing more when the E-field is along the shortest bond in the basal plane. This anisotropy is an indication of the crucial rôle played by the SOC in the spin dissipation. Indeed, switching off the SOC (setting $\alpha = 0$ in Eq. (16)) results in net spin conservation during and following the optical excitation.

A further observation is that clusters of different sizes demagnetize at different rates for the same pulse, as shown in Fig. 3. In fact, the $Fe_2$, $Fe_4$, and $Fe_6$ clusters investigated also represent atomic structures of increasing dimensionality: going from linear ($Fe_2$) to planar ($Fe_4$) to full three dimensional ($Fe_6$). All structures, simulated in large ideally-reflecting cuboid simulation boxes, show ultrafast demagnetization, namely, a decrease of the total spin, $S_{tot}^z(t) = \int d^3r\, m^z(\mathbf{r}, t)$, when excited by a fs E-field pulse along a non-zero dimension (i.e., for $Fe_2$ along the bond and for $Fe_4$ in the cluster plane). Interestingly, the long-term rate of demagnetization increases with the cluster size (see Fig. 3a) which is suggestive of a nonlocal mechanism (see Eq. 15). However, in the early times after the pulse, the TDSDFT simulations show very similar relative demagnetization rate in all three clusters. This is an indication of a common demagnetization onset mechanism, and it corroborates the key rôle played by the SOC close to the Fe nuclei.

A spatial analysis of the spin density evolution shows that after the pulse, nearly all the spin drop takes place in the vicinity of the atomic centers (see Fig. 3c, d) for
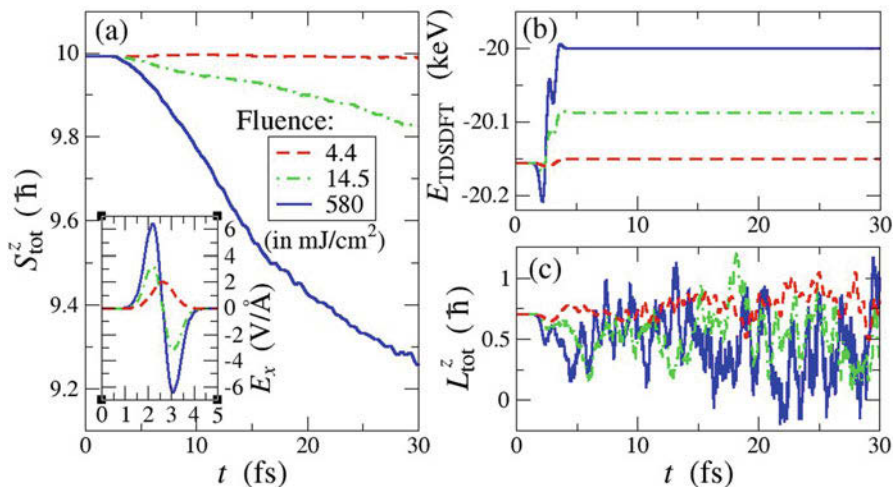
**Fig. 4** Time evolution of (**a**) the total spin, (**b**) the total TDSDFT energy, and (**c**) the total KS orbital momentum, defined as $L_{tot}^z(t) = \int d^3r\, L_{KS}^z(\mathbf{r}, t)$ for the $Fe_6$ cluster excited by pulses of different shapes (see inset of panel (**a**) with the corresponding color code). (Figure adapted from Stamenova et al. 2016)

both $Fe_4$ and $Fe_6$. We define local atom-based spin as the integral of the spin density, $\mathbf{m}(\mathbf{r}, t)$, inside a sphere $\Sigma_i$ centered on the $i$-th nucleus, $\mathbf{S}_{loc}^i(t) = \int_{\Sigma_i} d^3r\, \mathbf{m}(\mathbf{r}, t)$. The total local spin is obtained by summing up over all atomic centers $i$. During the pulse it appears that spin and charge rapidly spill out from the atomic centers and occupy the interatomic regions. This is not accompanied by a total (net) spin variation as practically no spin flipping occurs during this ultrafast excitation stage. After this the spin starts to decay, and all the SOC-driven spin-flip events take place in the vicinity of the atomic cores. The bipolar pulse excites a coherent antiphase oscillation of the local spins along the direction of the pulse, which persists as all local spins continue to decay (see Fig. 3e, f). The very high frequency oscillations of the local spin quantities in Fig. 3 are due to sloshing spin currents and include some minor numerical noise from the spatial integration.

The demagnetization rate depends strongly also on the shape and amplitude of the pulse, as shown in Fig. 4. A small unipolar pulse is shown to generate very little demagnetization (even exhibiting time intervals of small total momentum enhancement), while bipolar pulses of increasing amplitude tend to demagnetize the cluster more. After the pulse, the total TDSDFT energy reaches a new higher value, which is kept constant for the rest of the simulation. This means that the cluster remains in an excited state, since there is no energy dissipation mechanism in the method. However, unlike the total energy, the total angular momentum is not a constant of motion after the pulse. Figure 4d shows the rapidly oscillating total orbital momentum of the cluster, defined as $\mathbf{L}_{tot} = \int d^3r\, \mathbf{r} \times \mathbf{p}$. This does not mirror the smooth decay of the total spin. The reason for that lies in the

rotational symmetry breaking due to the superimposed ionic potentials of the six frozen atom cores. What is seen after the pulse is a transfer of spin into the orbital angular momentum, which in turn is not conserved and develops high frequency oscillations due to the elastic scattering of the excited electrons onto the immobile atomic cores. Note that the lack of total angular momentum conservation of the valence electron subsystem is physical and it stems from their interaction with the lattice. The TDSDFT simulations are aimed at describing the early stages of the demagnetization process, when other relaxation mechanisms (e.g., phonons) can be ignored.

## 4.2 Hydrodynamical Formulation of the Spin Continuity Equation

In order to understand the dynamics better, it is useful to rewrite the continuity equation, Eq. (14), in terms of the following set of hydrodynamical variables:

$$\boldsymbol{\mu}(\mathbf{r}, t) \equiv \frac{\mathbf{m}(\mathbf{r}, t)}{n(\mathbf{r}, t)} \qquad \text{and} \qquad \mathbf{v}(\mathbf{r}, t) \equiv \frac{\mathbf{j}_\mathrm{p}(\mathbf{r}, t)}{n(\mathbf{r}, t)}, \qquad (17)$$

leading to:

$$\frac{D}{Dt}\mathbf{m}(\mathbf{r}, t) + \sum_i \nabla \cdot \mathbf{v}_i(\mathbf{r}, t)\mathbf{m}_i(\mathbf{r}, t) =$$

$$= \nabla \cdot \mathscr{D}(\mathbf{r}, t) + \frac{1}{2c}\mathbf{m}(\mathbf{r}, t) \times \mathbf{B}_\mathrm{eff}(\mathbf{r}, t) + \frac{1}{4c^2}\langle (\nabla v_\mathrm{s}(\mathbf{r}, t) \times -i\nabla) \times \hat{\boldsymbol{\sigma}}\rangle \quad (18)$$

where $\mathbf{v}_i(\mathbf{r}, t)$ and $\mathbf{m}_i(\mathbf{r}, t)$ are the velocity field and the magnetization density associated to the single KS state $\psi_i^\mathrm{KS}(\mathbf{r}, t)$ and $\nabla \cdot \mathscr{D}(\mathbf{r}, t)$ is a new spin current flux term. Now, in addition to the spin-orbit, we have a torque due to the effective magnetic field:

$$\mathbf{B}_\mathrm{eff}(\mathbf{r}, t) = \mathbf{B}_\mathrm{s}(\mathbf{r}, t) + \frac{1}{\bar{\mathscr{F}}}\left[\frac{\nabla n \cdot \nabla \boldsymbol{\mu}}{n} + \nabla^2 \boldsymbol{\mu}\right], \qquad (19)$$

where $\bar{\mathscr{F}} = \frac{\langle \psi_i^\mathrm{KS}(\mathbf{r}, t)^* \psi_i^\mathrm{KS}\rangle}{n(\mathbf{r}, t)}$ defines an average occupation number of the KS states. This effective field is responsible for generating a torque upon the magnetization vector as we will see in the next sections. Furthermore it has been shown (see Simoni et al. 2017 for details) that the spin current divergence also contributes to the magnetization torque following an ultrafast optical excitation. It locally increases spin non-collinearity and enables the SOC to dissipate spins more effectively in the orbital momentum channel. Note that globally the total spin deflection from the quantization axis, $z$, in these simulations is negligible.

### 4.3   Identifying the Role of the Spin-Orbit Interaction

In order to quantify the effect of the SOC in the early stage of the demagnetization process, we have investigated the spin dynamics for different values of the scaling factor $\alpha$ (see Eq. (16)). Depicted in Fig. 5 is the effect on the global spin variation trajectory from tuning $\alpha$ between 0 and 4. Clearly, the spin loss rate is strongly affected by the SOC strength with the limit of $\alpha = 0$ (no SOC), resulting in global spin conservation. In panel (b) we have plotted the same spin trajectories after removing their GS offset and scaling them by a factor $1/\alpha^2$. The overlap of the curves demonstrates that in the initial coherent stage, the spin-decay rate scales as the square of the SOC strength.

Since a SOC scaling factor is just a theoretical tool, a number of different atomic species have been also considered. In response to the same laser pulse, clusters geometrically identical to $Fe_6$ (Gutsev and Beuschlicher 2003) but composed of Co and Ni (although this is not their GS geometry, it is chosen as to exclude structural factors from the comparison.) demagnetize at different rates, as shown in Fig. 5. We quantify their ionic SOC strength as

$$\lambda_{\text{eff}} = \sum_{l \in \text{occ.}} \frac{n_l}{n_{\text{tot}}} \int V_l^{\text{SO}}(\mathbf{r}) R_l^2(\mathbf{r}) d^3 r \;, \tag{20}$$
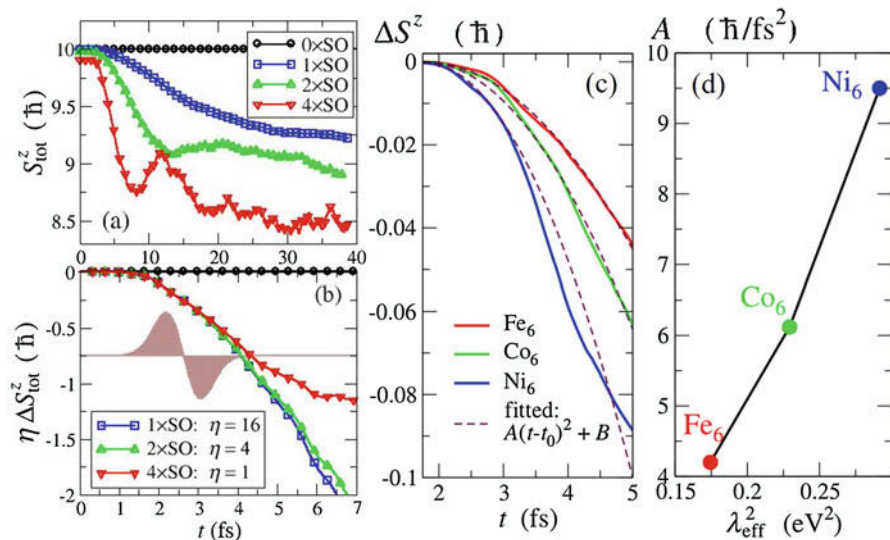


**Fig. 5** (a) Evolution of global spin expectation value for different values of the factor $\alpha = 0, 1, 2, 4$ scaling the SOC term in Eq. (16). (b) Same as in panel (a) but having the GS spin subtracted and a multiplication factor $\eta = (\alpha_{\text{max}}/\alpha)^2$. (c) TDSDFT trajectories of the variation of $S_{\text{tot}}^z(t)$ with respect of the GS spin for three different clusters: $Fe_6$, $Ni_6$, and $Co_6$ (all sharing the same $Fe_6$ geometry) fitted to $y = A(t - t_0)^2 + A_0$ (dashed curves). (d) The fitting coefficients $A$ from panel (c) versus the effective atomic SOC from Eq. (20) for each material. (Figure adapted from Stamenova et al. 2016)

where $n_l$ are the KS state occupations with $l$ spanning the valence states (in this case $3s$, $3p$, $3d$, and $4s$), $n_{\text{tot}} = \sum_{l \in \text{occ.}} n_l$, $R_l(\mathbf{r})$ are the radial pseudo-atomic wavefunctions, and $V_l^{\text{SO}}(\mathbf{r}) = \frac{2l}{2l+1} \left[ V_{\text{PP}}^{l+1/2}(\mathbf{r}) - V_{\text{PP}}^{l-1/2}(\mathbf{r}) \right]$ is the same object as in Eq. 16. Those are all obtained from LSDA calculations for isolated atoms. By fitting the first few femtoseconds of the demagnetization curve to a quadratic time decay $S^z(t) \propto A(t - t_0)^2$ with $t_0$ in the rise of the laser pulse (around 1.5 fs), we extract the demagnetization rate $A$ for each of the three different clusters. Such extracted demagnetization rates show a systematic (quasi-linear) dependence on their ionic SOC strength $\lambda_{\text{eff}}^2$ (see Fig. 5c).

## 4.4    Simple Model

A minimal model involving SOC and resulting in a local variation of the spin expectation value is described by Stamenova et al. (2016):

$$\hat{H}(t) = \lambda \hat{\mathbf{L}} \cdot \hat{\mathbf{S}} + \mathbf{B}_{\text{eff}}(t) \cdot \hat{\mathbf{S}}, \tag{21}$$

where $\mathbf{B}_{\text{eff}}(t)$ represents a generic time-dependent magnetic field, $\hat{\mathbf{S}}$ is the spin operator, and $\hat{\mathbf{L}}$ is the orbital momentum operator. This model is designed to mimic the local TDSDFT spin dynamics at a given spatial point, but it does not reproduce the full ab initio simulations in all their features.

By solving the model for a simple two-level system, we observe that a temporal change of the magnetic field, $\mathbf{B}_{\text{eff}}(t)$, combined with finite SOC leads to the transfer of part of the spin expectation value into the orbital momentum channel more efficiently than when $\mathbf{B}_{\text{eff}}$ is constant. The amount of spin lost after the action of $\mathbf{B}_{\text{eff}}(t)$ is not the same for different SOC coefficients $\lambda$, and, in particular, higher $\lambda$ values lead to faster momentum transfers, i.e., faster spin decays.

A numerical integration of the Schrödinger equation with the Hamiltonian of Eq. (21) in the case of no SOC (i.e., $\lambda = 0$) returns, as expected, no observable spin dynamics, if the effective magnetic field $\mathbf{B}_{\text{eff}}(t)$ stays parallel to the spin expectation value $\langle \hat{\mathbf{S}} \rangle$. In contrast, when $\lambda \neq 0$ and the initial state has not the maximum or minimum $j_z = l_z + s_z$, a steplike variation of $\mathbf{B}_{\text{eff}}(t)$ produces a sharp variation in the spin expectation value $\langle \hat{S}_z \rangle$ (see Fig. 6). In particular, for an initial spin-up state ($s_z = 1/2$), we observe a decrease of $\langle \hat{S}_z \rangle$ after the drop of $|\mathbf{B}_{\text{eff}}(t)|$, while an initial spin-down state ($s_z = -1/2$) results in an increase of the spin expectation value.

This simple model demonstrates that a change in the magnetic field $\mathbf{B}_{\text{eff}}(t)$, combined with SOC, leads to a transfer of part of the spin expectation value into the orbital momentum more efficiently than when $\mathbf{B}_{\text{eff}}(t) = 0$. Figure 6b further reveals that the amount of spin lost after the action of $\mathbf{B}_{\text{eff}}(t)$ depends strongly on the value of the SOC coefficient $\lambda$. In particular, larger values of $\lambda$ lead to faster spin-decay rates. Furthermore, it can be established numerically (and analytically, see Stamenova et al. (2016)) that there is a quadratic dependence of the spin-decay rate on the SOC strength (see Fig. 6c).
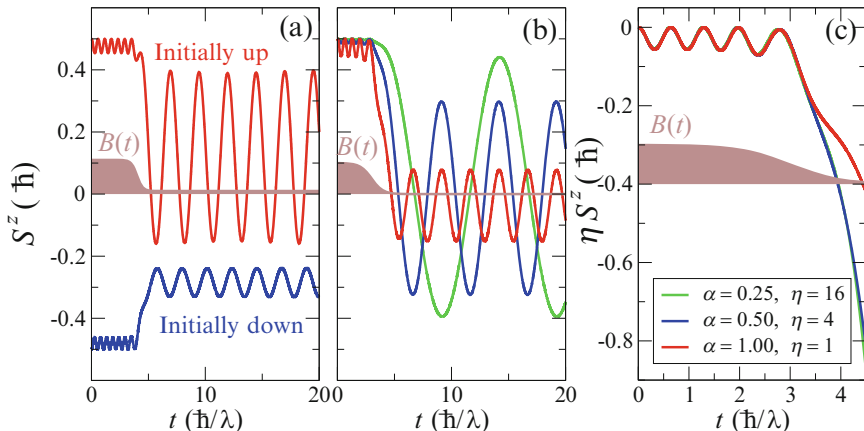
**Fig. 6** Trajectories of $\left\langle \hat{S}_z \right\rangle$ under the action of the Hamiltonian of Eq. (21) in the case of (**a**) having an initial spin-up or spin-down state, (**b**) for different values of a multiplicative factor $\alpha$ to the SOC ($\lambda \rightarrow \alpha\lambda$) and for an initial spin-up state, (**c**) the trajectories rescaled by a factor $1/\alpha^2$. The shaded area describes the temporal profile of the applied magnetic field $[0, 0, B_{\text{eff}}(t)]$. The initial orbital configuration is a linear combination of $l_z = 0, 1$ states. (Figure adapted from Stamenova et al. 2016)

The insights drawn from the simple model map out a possible interpretation for the ultrafast demagnetization observed in the TDSDFT simulations. An effective magnetic field provides the connection between the light pulse and the spin dynamics. This is $\mathbf{B}_{\text{eff}}(\mathbf{r}, t)$, identified in Eq. (19). Such field is directly coupled to the spin density and dependents on the density gradients and hence couples the spins with the laser-induced charge dynamics. In order to show that $\mathbf{B}_{\text{eff}}(\mathbf{r}, t)$ initiates the spin demagnetization, as suggested by the simple model, we compare the variation of the spin density in the $(x-y)$ plane for the $Fe_6$ cluster averaged over the simulation time and the axial direction (see Fig. 7). Together with highlighting the localized character of the demagnetization effect in small areas close to nuclei, Fig. 7 also shows the spatial correlation between regions of high spin loss with regions of high $\mathbf{B}_{\text{eff}}$ variation.

## 5    Ultrafast Spin Dynamics in Bulk Materials

TDSDFT simulations may also be performed for bulk systems using the formulation of Eq. (13), which has been implemented in the all-electron electronic structure code ELK `elk.sourceforge.net` for *two*-component Pauli spinors, allowing systems with non-collinear magnetism to be simulated.

The ELK basis set consists of linearized augmented plane waves (LAPW). These decompose the system into so-called muffin-tin regions centered on each atom and the interstitial region (IR) in between them. The calculations discussed below used
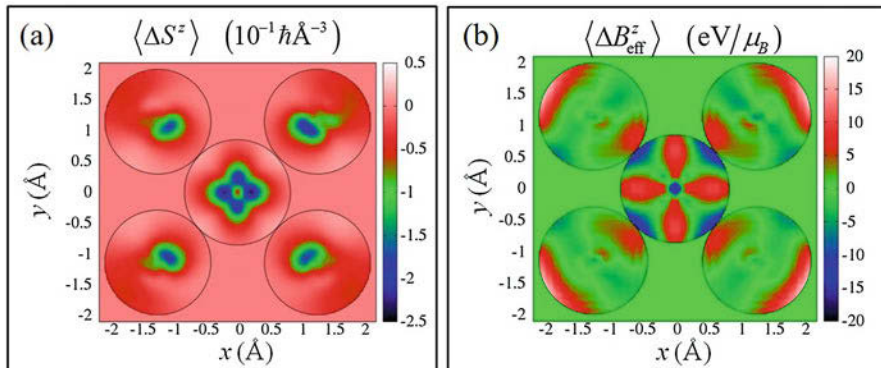
**Fig. 7** Contour plots of computed observables averaged in time (entire simulation of 30 fs) and along the $z$-direction (perpendicular to the page and only within the marked spheres of radius $0.85$ Å around each atom). Here we present the variation from the GS values for (**a**) the spin density $S^z(\mathbf{r})$ and (**b**) the effective magnetic field $B_{\text{eff}}^z(t)$. (Figure adapted from Stamenova et al. 2016)

a $8 \times 8 \times 8$ $k$-points grid and a timestep $\Delta t = 0.1$ au. Further computational details can be found in references Krieger et al. (2015); Elliott et al. (2016a, b). In all cases we begin the simulation from the GS with the global spin moment parallel to the $z$-axis.

We will discuss in detail two forms of spin dynamics observed in bulk systems, namely, ultrafast demagnetization and ultrafast spin transfer between sublattices.

## 5.1 Ultrafast Demagnetization of Bulk Materials

The first experimental observation of ultrafast demagnetization consisted in measuring the rotation of the light polarization reflected by bulk Ni [magneto-optical Kerr effect (MOKE)] (Beaurepaire et al. 1996; Hohlfeld et al. 1997; Scholl et al. 1997; Koopmans et al. 2000). Such change in the MOKE signal, induced by a strong laser pulse, was found to take place in under 100 fs. Since then, subsequent experiments have found changes as fast as 10 s of fs and have been observed also in bulk Fe (Carpene et al. 2008; La-O-Vorakiat et al. 2009), Co (Cinchetti et al. 2006), and Gd (Lisowski et al. 2005; Eschenlohr et al. 2014). Thus, it is natural to ask what spin dynamics are seen in TDDFT simulations for these systems.

In Fig. 8 we show the relative change in the global spin moment, $M(t)$, for bulk Fe, Ni, and Co following the application of a laser pulse (frequency 2.72 eV, FWHM 2.5 fs, peak intensity $10^{15}$ W/cm$^2$), also shown in the figure. The same simulations run in the absence of SOC return no demagnetization, indicating once again that SOC is the interaction responsible for the loss of global spin moment. This confirms the analytic result derived in Eq. (15). Note that in Fig. 8 we plot the absolute value $M(t) = |\mathbf{M}(t)|$, which rules out any canting of the moment that leaves $M(t)$ unchanged but reduces $\mathbf{M}_z(t)$.
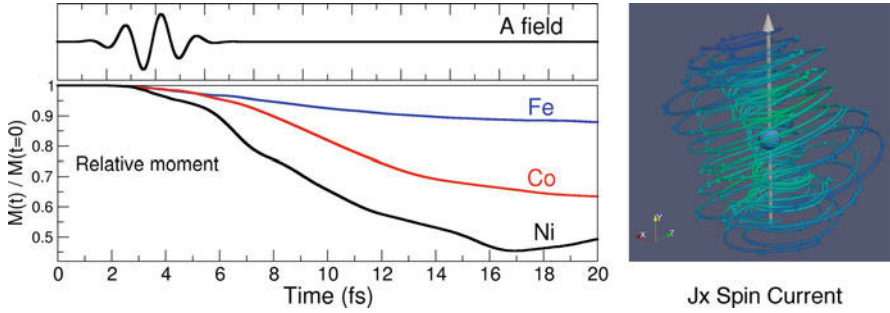
**Fig. 8** Demagnetization of bulk Ni, Fe, and Co. In the left-hand side plot, we show the total magnetic moment $M(t) = |\mathbf{M}(t)|$ as a function of time, while in the right-hand side one, we trace in space the vector field position of the spin current, $\mathbf{j}_x$. (Figure adapted from Krieger et al. (2015) and Krieger et al. 2017)

Another interesting feature that may be seen in Fig. 8 is the separation of timescales between the laser excitation and the demagnetization process. During the laser pulse, electrons are excited into higher energy states, but the global spin moment does not change significantly. It is only in the $\approx 10 - 20$ fs following the laser pulse peak that the demagnetization takes place. Only by utilizing such a short laser pulse are we able to disentangle these two physical phenomena. For pulses of longer duration, the two occur together. We will now study the EOM, Eq. (15), for the global moment in order to explain why this time lag is present and what causes the demagnetization.

Focusing on the $z$-component of Eq. (15):

$$\frac{\partial}{\partial t} M_z(t) = \frac{1}{2c^2} \int d^3r \left[\nabla v_s(\mathbf{r}, t) \times \mathbf{j}_y(\mathbf{r}, t)\right]_x - \left[\nabla v_s(\mathbf{r}, t) \times \mathbf{j}_x(\mathbf{r}, t)\right]_y , \quad (22)$$

where $\mathbf{j}_\mu$ is the spin current density, defined previously, which describes the flow of the $\mu$ spin component throughout the system. We see that the demagnetization can only occur when the $\mathbf{j}_x$ spin current circulates about the $y$-axis and/or $\mathbf{j}_y$ circulates around the $x$-axis. In Fig. 8, we plot the streamlines found by following the vector field of $\mathbf{j}_x$. This allows us to visualize the flow of spin current at a particular time during the demagnetization process. Indeed we see exactly the circulation required by Eq. (22). From this analysis, we understand that the role of the laser pulse is to excite the system (including changing the exchange coupling) and to cause such spin currents to appear (note similar spin currents exist in the GS due to the spin texture induced by the SOC). The SOC can then rotate the spin components, causing a decrease in $M_z(t)$ leading to the demagnetization observed. These processes operate on different timescales, primarily dictated by the strength of the SOC, the exchange field, and the laser pulse duration.

Ultrafast demagnetization experiments are generally performed with the ferromagnetic material grown on top of a conducting substrate. This allows spin current
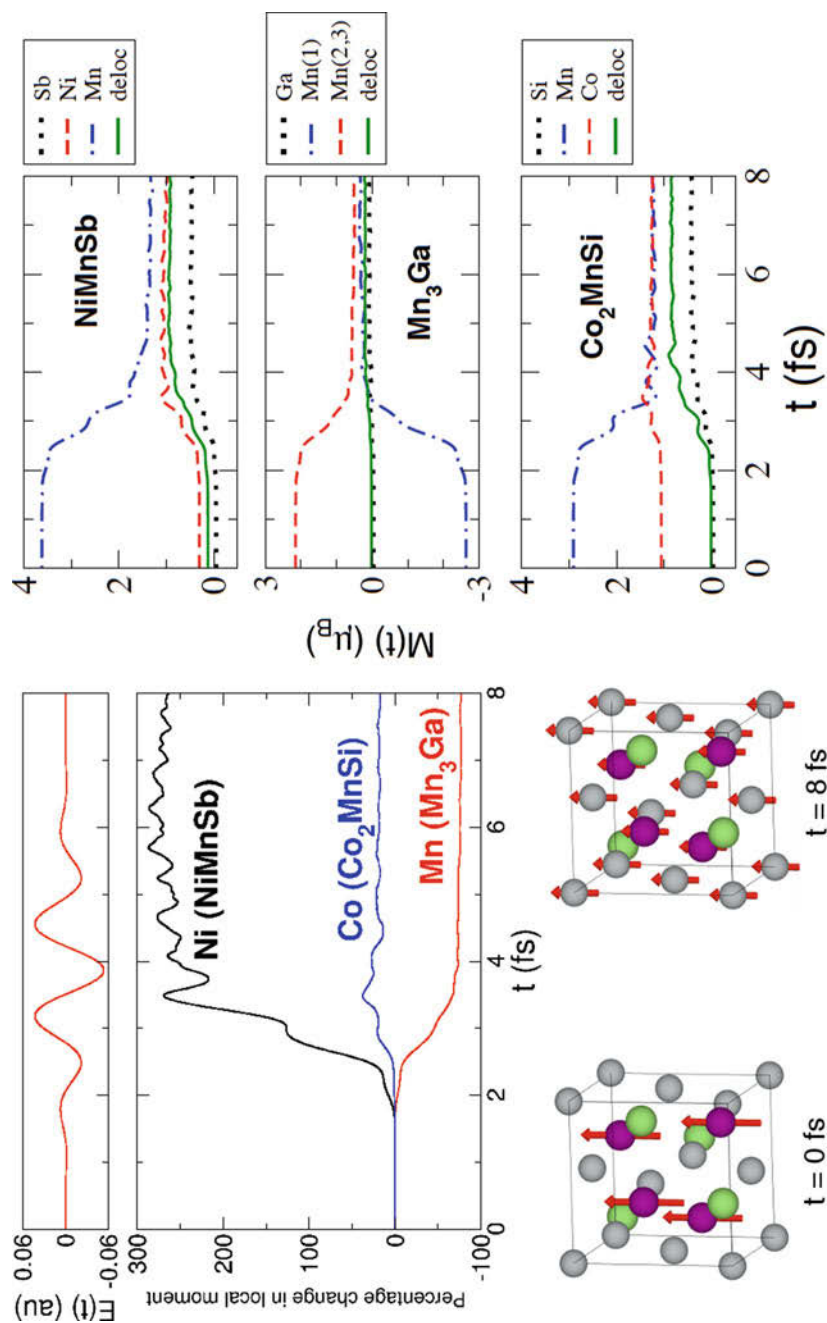
**Fig. 9** (continued)

to transport moment away from the probed region, also leading to an ultrafast demagnetization of this region, as in the super-diffusive spin-transport model (Battiato et al. 2010). Due to computational restrictions on the size of supercell possible, it is difficult to include this form of demagnetization in our TDSDFT simulations. However, recent experiments (Razdolski et al. 2017) used an insulating substrate to eliminate any super-diffusive mechanism, yet ultrafast demagnetization of Ni, Fe, and Co was still observed, suggesting that the SOC-orbit mediated spin-flips observed in our simulations have been experimentally observed.

Finally, we should comment on the fact that we purposefully used an ultrashort, ultrastrong laser for the simulations presented. As discussed, this is to disentangle the many physical processes involved in ultrafast demagnetization. However, if we use laser parameters taken from experiments, we still observe significant demagnetization, as can be seen in Krieger et al. (2017).

## 5.2   Optical Inter-sublattice Spin Transfer

The magnetization dynamics studied so far have all involved a decrease in the global spin-magnetic moment driven by SOC. However, another interesting phenomenon seen in TDSDFT calculations for systems with more than one magnetic sublattice is an increase of the local moment, driven by spin currents transferring angular momentum from one sublattice to another (Elliott et al. 2016b). This form of magnetization dynamics can have a dramatic effect on the local magnetic moments, as can be seen in Fig. 9. Here we plot the percentage change in the local spin moment (defined as the integral of the magnetization density inside an appropriately sized sphere around each atom) for a particular atom (Ni, Co, Mn) in three Heusler alloys, NiMnSb, $Co_2MnSi$, and $Mn_3Ga$. It is clear that the local moment reacts to the incoming laser pulse completely different in each of the three cases: it may be strongly enhanced (Ni), relatively unchanged (Co), or completely destroyed (Mn). Having the ability to selectively control these local magnetic moments could be of great technological importance, particularly for magnetic storage and spintronic devices. We now investigate the underlying physical mechanism for such inter-sublattice spin transfer, with the goal of identifying a strategy for tailoring the material and/or the pulse for a particular purpose.

Heusler alloys are compounds consisting of four interpenetrating $fcc$ sublattices with stoichiometry $X_2YZ$ (or $XYZ$ in the case of half-Heuslers), where $X$ and $Y$

**Fig. 9** Laser-induced magnetization dynamics in Heusler alloys. The left-hand side panel presents the relative change in local magnetic moment of the given atom within NiMnSb, $Co_2MnSi$, and $Mn_3Ga$ and the shape of the electric field pulse used to initiate the dynamics. We also show two snapshots of the local spin configuration of NiMnSb at $t = 0$ (ground state) and $t = 8$ fs. Mn atoms are in purple, Ni in green, and Sb in gray. The exact values can be deduced from the right-hand side panels which display the dynamics of the local moment and of the interstitial region ("deloc") for the three studied compounds. (Figure adapted from Elliott et al. 2016b)

are usually transition metals and $Z$ is a main group element. We choose to illustrate the inter-sublattice spin transfer dynamics using Heusler alloys as they exhibit a wide range of magnetic behavior and thus can be a convenient playground for understanding the phenomenon. As already mentioned, the three compounds we will focus on are NiMnSb, $Co_2MnSi$, and $Mn_3Ga$. The first two are ferromagnets, while $Mn_3Ga$ is a ferrimagnet with antiferromagnetic coupling between the two different Mn sublattices. For more information on the crystal structure of these compounds, see Elliott et al. (2016b) and references therein.

The local spin dynamics for each atom and for the interstitial region (shown in Fig. 9) reveal why an increase or a reduction of the local magnetization was observed. Starting with NiMnSb, the main features to observe are a large Mn moment loss and an increase in moment for both Ni and the IR. Due to the small initial moment on Ni, an increase of $0.7\mu_B/f.u.$ due to the laser pulse manifests itself as a huge change in the percentage moment seen in Fig. 9. A similar picture emerges for $Co_2MnSi$ with a loss of Mn moment and a gain in the IR moment. However in contrast to NiMnSb, there is not a large increase in Co moment. Finally, in $Mn_3Ga$ there are two symmetry positions that the Mn atoms can occupy, labeled as Mn(1) (Wyckoff position 2b of the $I4/mmm$ symmetry group) and Mn(2,3) (4d). In the ground state, Mn(1) and Mn(2,3) are coupled antiferromagnetically. The laser pulse then strongly decreases the moment on each Mn atom. In fact, for Mn(1) this is completely destroyed, although for less intense laser fields, it simply decreases but does not vanish. Note that the global spin moment during this time remains unchanged. For longer times, the spin-orbit mediated demagnetization discussed previously may also take place. Again by using an ultrashort laser pulse (frequency 2.72 eV, FWHM 2.42 fs, peak intensity $1 \times 10^{14}$ W/cm$^2$) in these simulations, we were able to separate the SOC-induced effects from purely optical effects, which we wish to focus on. From these three examples, it is clear that the laser pulse induces spin moment to flow from one sublattice to another, a phenomenon that we term as *optical inter-sublattice spin transfer*.

We can make use of the TDSDFT simulation to understand in more detail the inter-sublattice spin transfer process. In particular, we can distinguish between majority and minority spin transfer, information not available from the magnetization density alone. For example, consider the transfer of moment between Ni and Mn in NiMnSb. We observed that the Ni moment increases while that of Mn decreases. This could be explained as a transfer of majority spins from Mn to Ni. However, it can also be explained as a transfer of minority spin from the Ni to Mn. In order to distinguish those processes, we plot in Fig. 10 the change in the number of majority and minority spins for each atomic specie, as defined by:

$$\Delta N_{\uparrow\downarrow}(t) = \frac{1}{2}\left(\Delta N(t) \pm \Delta M_z(t)\right) , \tag{23}$$

where $\Delta N(t)$ is the change in local charge and $\Delta M_z(t)$ is the change in the $z$-component of the local magnetic moment (the $z$-direction is that of the ground-state moment). If the system remains approximately collinear during the laser pulse
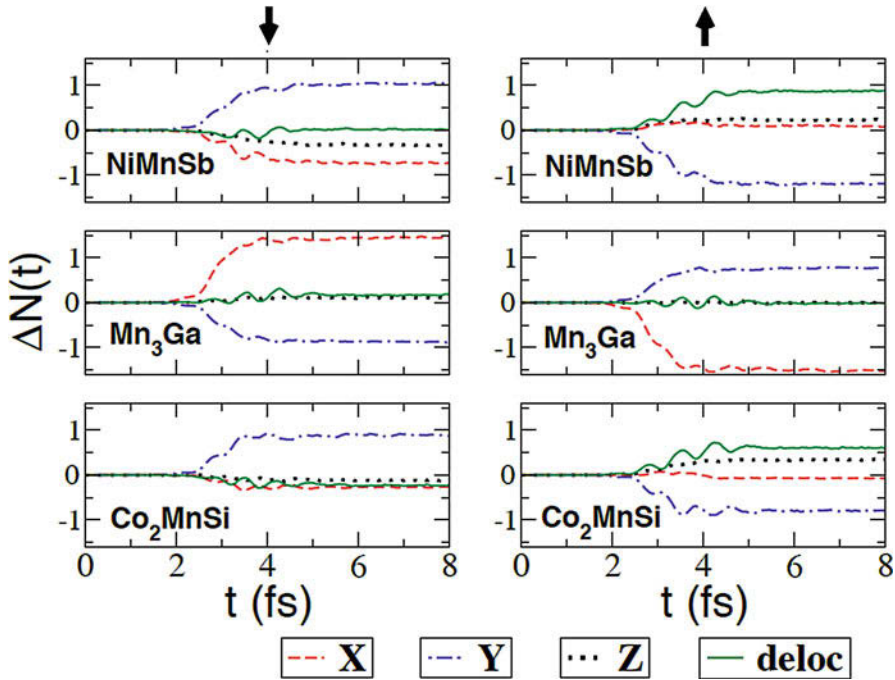
**Fig. 10** The change in majority (↑) and minority (↓) local population for each atoms of the Heusler alloys considered. The spin direction is defined from the ground-state magnetic structure. The labels of the different atoms follow the general formula unit definition $X_2YZ$ ($XYZ$ for half-Heusler alloys)

and no spin-flip transitions take place, then the sum of $\Delta N_{\uparrow\downarrow}(t)$ over all atoms (and the IR) will be zero (i.e., the total number of majority/minority electrons is conserved). Again starting with NiMnSb, we see a gain of minority electrons on Mn, which is accompanied by a loss of minority electrons on Ni. This is a transfer of minority electrons from Ni to Mn. Thus, we find the perhaps counter-intuitive answer of how the Ni local moment increases. In the majority channel, we see a further loss of Mn moment due to transfer of majority electrons from the Mn to the IR. This picture of inter-sublattice spin transfer is further validated by the cases of $Co_2MnSi$ and $Mn_3Ga$. The Mn atom in $Co_2MnSi$ behaves similarly as in NiMnSb, in that it gains minority and losses majority electrons, leading to a loss of its moment. However, now Co loses a similar amount of majority and minority electrons, leaving its moment unchanged. In $Mn_3Ga$, Mn(1) and Mn(2,3) behave opposite to each other, with Mn(1) transferring majority electrons to the two Mn(2,3) atoms and the two Mn(2,3) atoms transferring minority electrons to Mn(1). This causes the local moment to be reduced in magnitude at all Mn sites.

In quantum mechanics we are used to the idea of electrons transitioning from an occupied state to an unoccupied one, due to the application of a laser pulse. When
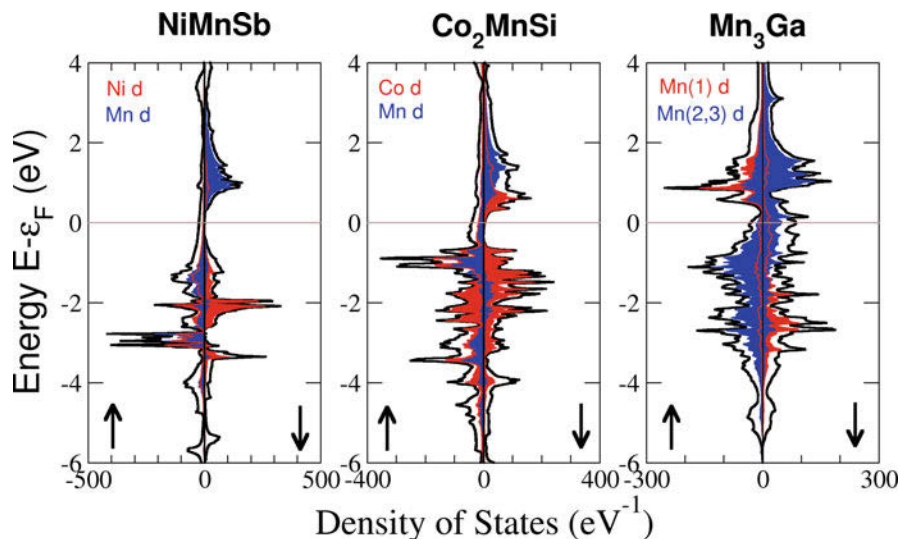
**Fig. 11** The ground-state DOS for NiMnSb, $Co_2FeSi$, and $Mn_3Ga$. The solid line is the total DOS, while the colored areas are the contributions from the $d$ shell of the various magnetic transition metals

considering magnets with different magnetic sublattices, for each spin channel, we can think of transitions from states primarily located on one atom to states primarily located on another, similar to the charge-transfer processes of molecular chemistry. Therefore, in order to determine what transitions are possible, it is useful to consider the GS density of states (DOS). This can only be a zero-th order approximation to the transition energies and states, as it is known from linear-response TDDFT (e.g., Sharma et al. 2014) that the KS response is corrected to the true response via the Hartree-XC kernel. Furthermore, the DOS does not tell us the probability of a particular transition. Keeping this in mind, we now ask if these inter-sublattice transitions can be understood from the DOS.

In Fig. 11 we plot the total DOS and the $d$-state atom-projected DOS for the three systems under investigation. As would be expected for magnets, we see in all cases minority $d$ bands above the Fermi energy, $\varepsilon_F$. This leads to a significant unoccupied DOS, which is available to the laser-excited electrons. For example, Mn in NiMnSb has a large moment of $3.62\,\mu_B$ in our GS DFT calculations, which leads to a large density of unoccupied Mn $d$-states centered approximately 1 eV above $\varepsilon_F$. Ni, by contrast, has no strong features above $\varepsilon_F$ but does present occupied $d$-states from 0 to 3 eV below $\varepsilon_F$ in both spin channels. Thus, in this simple picture of transitions from occupied to unoccupied state, we would anticipate minority spin transfer from Ni to Mn, as it is the transition with the highest density of both unoccupied and occupied states. This is precisely the transition we have observed in our simulations, demonstrating the usefulness of the DOS for interpretation of results.

Clearly, we observe several other features in our time-dependent calculations that lie outside the DOS picture due to many-body and nonlinear effects, but the DOS still provides an initial interpretation tool. This is confirmed by looking at $Co_2MnSi$ and $Mn_3Ga$ in Fig. 11. The situation for $Co_2MnSi$ is very similar to that of NiMnSb except that there is a large number of unoccupied Co minority states above $\varepsilon_F$. This reduces the net amount of minority electrons transferred from Co to Mn, reducing any increase in the Co moment. The antiferromagnetic coupling in $Mn_3Ga$ makes the DOS interpretation quite simple. On Mn(1) there are occupied majority states and unoccupied minority ones ($d^5$ configuration), while Mn(2) and Mn(3) have fractional occupation for both majority and minority spins. The laser then excites majority electrons from Mn(1) to Mn(2,3) and minority from Mn(2,3) to Mn(1), decreasing the amount of moment on each ion. In Elliott et al. (2016b), the dependence on the laser intensity and frequency was investigated, and the results were found to be consistent with the DOS picture. In particular, the frequency dependence can be inferred by observing the energy difference between the bulk of the occupied and unoccupied DOS.

The inter-sublattice spin transfer observed in our TDSDFT simulations can now be explained using the GS DOS, which provides a simple picture for this phenomenon. Although we choose to illustrate this mechanism in Heusler compounds, the DOS interpretation suggests that this is general to any system with inequivalent magnetic sublattices. The strength and frequency of the transitions depend on the material details and on the laser intensity (as nonlinear processes likely contribute). Most importantly our predictions can be validated experimentally by using element-specific techniques such as X-ray magnetic circular dichroism or XUV TR-MOKE.

## 6    Conclusion

In this chapter we have shown the most recent advances in the study of laser-induced magnetization dynamics in both magnetic clusters and periodic solids, showing different magnetic orders. In particular we have focused our attention on the early dynamical evolution, namely, on the first few tens of femtoseconds following the laser excitation. This is the timescale where the dynamics is entirely driven by the electronic degrees of freedom and where the magnetic moment, the exchange coupling, and the spin-orbit interaction are all fast-varying time-dependent quantities. As such it is a regime in which models that do not explicitly account for the electrons are incapable to make predictions.

The theory of choice in this limit is time-dependent density functional theory. TDDFT has a rigorous foundation, which is reviewed here, and represents the only viable approach to the dynamics. In fact, it is fully quantitative and parameter free, but at the same time, it is computationally light enough to guarantee long simulations for real materials. In general we have shown that the application of an intense laser field is capable of producing ultrafast demagnetization. This, however, takes place only in presence of spin-orbit interaction and when the system investigated is able to sustain a local current (an atom in vacuum does not

demagnetize). The loss of magnetization, following the laser excitation, is the result of the angular momentum transfer between the spin and the orbital components. Firstly, part of the spin moment is transferred to the electronic angular momentum. This is a rapid process, whose timescale is set by the strength of the spin-orbit interaction. As a second step, the electronic angular momentum is dissipated to the angular momentum of the nuclei, which are described as stationary at our level of theory.

Being a fully quantitative theory, TDDFT allows us to investigate complex magnetic orders. For instance, we have discussed the case of antiferromagnetic metals and shown that the laser can produce inter-lattice spin transfer. This is an active channel for spin dynamics, which may result in an increase of one or more of the sublattice magnetizations. In conclusion we have shown that TDDFT may be a powerful tool for the study of ultrafast spin dynamics in complex magnetic systems. Although in its present form the theory is non-dissipative, namely, it cannot describe the long-time magnetization recovery, it still represents the only fully atomistic approach to spin dynamics available. We envision that TDDFT will have a crucial role in the understanding of complex spin phenomena, with applications to materials science, spintronics, and quantum computing.

# References

Battiato M, Carva K, Oppeneer PM (2010) Superdiffusive spin transport as a mechanism of ultrafast demagnetization. Phys Rev Lett 105:027203

Beaurepaire E, Merle J, Daunois A, Bigot J (1996) Ultrafast spin dynamics in ferromagnetic nickel. Phys Rev Lett 76(22):4250–4253

Boeglin C, Beaurepaire E, Halte V, Lopez-Flores V, Stamm C, Pontius N, Durr HA, Bigot JY (2010) Distinguishing the ultrafast dynamics of spin and orbital moments in solids. Nature 465:458–461

Capelle K, Vignale G, Gyorffy B (2001) Spin currents and spin dynamics in time-dependent density-functional theory. Phys Rev Lett 87(20):206403

Carpene E, Mancini E, Dallera C, Brenna M, Puppin E, De Silvestri S (2008) Dynamics of electron-magnon interaction and ultrafast demagnetization in thin iron films. Phys Rev B 78(17):174422

Castro A, Appel H, Oliveira M, Rozzi CA, Andrade X, Lorenzen F, Marques MAL, Gross EKU, Rubio A (2006) Octopus: a tool for the application of time-dependent density functional theory. Phys Stat Sol B 243:2465

Cinchetti M, Albaneda MS, Hoffmann D, Roth T, Wuestenberg JP, Krauss M, Andreyev O, Schneider HC, Bauer M, Aeschlimann M (2006) Spin-flip processes and ultrafast magnetization dynamics in co: unifying the microscopic and macroscopic view of femtosecond magnetism. Phys Rev Lett 97(17):177201

Elliott P, Mueller T, Dewhurst JK, Sharma S, Gross EKU (2016a) Ultrafast laser induced local magnetization dynamics in heusler compounds. Sci Rep 6:38911

Elliott P, Krieger K, Dewhurst JK, Sharma S, Gross EKU (2016b) Optimal control of laser-induced spin–orbit mediated ultrafast demagnetization. New J Phys 18:013014. https://doi.org/10.1088/1367-2630/18/1/013014

Eschenlohr A, Sultan M, Melnikov A, Bergeard N, Wieczorek J, Kachel T, Stamm C, Bovensiepen U (2014) Role of spin-lattice coupling in the ultrafast demagnetization of $Gd_{1-x}Tb_x$ alloys. Phys Rev B 89(21):214423

Gutsev GL, Beuschlicher CW (2003) Electron affinities, ionisation energies, and fragmentation energies of $fe_n$ clusters ($n = 2-6$): a density functional theory study. J Phys Chem 107:7013–7023

Hinschberger Y, Hervieux PA (2012) Foldy wouthuysen transformation applied to the interaction of an electron with ultrafast electromagnetic fields. Phys Lett A 376:813

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864

Hohlfeld J, Matthias E, Knorren R, Bennemann K (1997) Nonequilibrium magnetization dynamics of nickel. Phys Rev Lett 78(25):4861–4864

Jean-Yves Bigot MV, Beaurepaire E (2009) Coherent ultrafast magnetism induced by femtosecond laser pulses. Nat Phys 5:515

Kampfrath T, Sell A, Klatt G, Pashkin A, Maehrlein S, Dekorsy T, Wolf M, Fiebig M, Leitenstorfer A, Huber R (2011) Coherent terahertz control of antiferromagnetic spin waves. Nat Photonics 5(1):31–34

Kazantseva N, Hinzke D, Nowak U, Chantrell RW, Atxitia U, Chubykalo-Fesenko O (2008) Towards multiscale modeling of magnetic materials: simulations of FePt. Phys Rev B 77(18):184428

Koopmans B, van Kampen M, Kohlhepp J, de Jonge W (2000) Ultrafast magneto-optics in nickel: magnetism or optics? Phys Rev Lett 85(4):844–847

Koopmans B, Malinowski G, Dalla Longa F, Steiauf D, Fahnle M, Roth T, Cinchetti M, Aeschlimann M (2010) Explaining the paradoxical diversity of ultrafast laser-induced demagnetization. Nat Mater 9:259–265

Krieger K, Dewhurst JK, Elliott P, Sharma S, Gross EKU (2015) Laser-induced demagnetization at ultrashort time scales: predictions of TDDFT. J Chem Theory Comput 11(10):4870–4874

Krieger K, Elliott P, Müller T, Singh N, Dewhurst JK, Gross EKU, Sharma S (2017) Ultrafast demagnetization in bulk versus thin films: an ab initio study. J Phys Condens Matter 29(22):224001

Kubler J, Hock KH, Sticht J, Williams AR (1988) Density functional theory of non-collinear magnetism. J Phys F Met Phys 18:469

La-O-Vorakiat C, Siemens M, Murnane MM, Kapteyn HC, Mathias S, Aeschlimann M, Grychtol P, Adam R, Schneider CM, Shaw JM, Nembach H, Silva TJ (2009) Ultrafast demagnetization dynamics at the $m$ edges of magnetic elements observed using a tabletop high-harmonic soft x-ray source. Phys Rev Lett 103:257402

Li Y, Ullrich CA (2008) Time-dependent v-representability on lattice systems. J Chem Phys 129(4):44105

Lisowski M, Loukakos P, Melnikov A, Radu I, Ungureanu L, Wolf M, Bovensiepen U (2005) Femtosecond electron and spin dynamics in Gd(0001) studied by time-resolved photoemission and magneto-optics. Phys Rev Lett 95(13):137402

Ostler TA, Barker J, Evans RFL, Chantrell RW, Atxitia U, Chubykalo-Fesenko O, El Moussaoui S, Le Guyader L, Mengotti E, Heyderman LJ, Nolting F, Tsukamoto A, Itoh A, Afanasiev D, Ivanov BA, Kalashnikova AM, Vahaplar K, Mentink J, Kirilyuk A, Rasing T, Kimel AV (2012) Ultrafast heating as a sufficient stimulus for magnetization reversal in a ferrimagnet. Nat Commun 3:666

Razdolski I, Alekhin A, Martens U, Buerstel D, Diesing D, Muenzenberg M, Bovensiepen U, Melnikov A (2017) Analysis of the time-resolved magneto- optical kerr effect for ultrafast magnetization dynamics in ferromagnetic thin films. J Phys Condes Matter 29(17):174002

Ruggenthaler M, van Leeuwen R (2011) Global fixed-point proof of time-dependent density-functional theory. EPL 95:13001

Runge E, Gross EKU (1984) Density functional theory for time-dependent systems. Phys Rev Lett 52:997

Schellekens AJ, Verhoeven W, Vader TN, Koopmans B (2013) Investigating the contribution of superdiffusive transport to ultrafast demagnetization of ferromagnetic thin films. Appl Phys Lett 102:252408

Scholl A, Baumgarten L, Jacquemin R, Eberhardt W (1997) Ultrafast spin dynamics of ferromagnetic thin films observed by fs spin-resolved two-photon photoemission. Phys Rev Lett 79(25):5146–5149

Sharma S, Dewhurst JK, Gross EKU (2014) Optical response of extended systems using time-dependent density functional theory. Top Curr Chem 347:235–257

Simoni J, Stamenova M, Sanvito S (2017) Ultrafast demagnetizing fields from first principles. Phys Rev B 95(2):024412

Stamenova M, Simoni J, Sanvito S (2016) Role of spin-orbit interaction in the ultrafast demagnetization of small iron clusters. Phys Rev B 94(1):014423

Stamm C, Kachel T, Pontius N, Mitzner R, Quast T, Holldack K, Khan S, Lupulescu C, Aziz EF, Wietstruk M, Durr HA, Eberhardt W (2007) Femtosecond modification of electron localization and transfer of angular momentum in nickel. Nat Mater 6:740–743

van Leeuwen R (1999) Mapping from densities to potentials in time-dependent density-functional theory. Phys Rev Lett 82(19):3863–3866. https://doi.org/10.1103/PhysRevLett.82.3863

Vodungbo B, Gautier J, Lambert G, Sardinha AB, Lozano M, Sebban S, Ducousso M, Boutu W, Li K, Tudu B, Tortarolo M, Hawaldar R, Delaunay R, Lopez-Flores V, Arabski J, Boeglin C, Merdji H, Zeitoun P, Luening J (2012) Laser-induced ultrafast demagnetization in the presence of a nanoscale magnetic domain network. Nat Commun 3:999

Walowski J, Münzenberg M (2016) Perspective: ultrafast magnetism and THz spintronics. J Appl Phys 120(14):140901. https://doi.org/10.1063/1.4958846

# Landau-Lifshitz-Bloch Approach for Magnetization Dynamics Close to Phase Transition

# 41

Oksana Chubykalo-Fesenko and Pablo Nieves

## Contents

### Abstract

Micromagnetic modeling has recommended itself as a useful tool for the design of magnetic nanostructures in multiple applications. The standard micromagnetics based on the integration of the Landau-Lifshitz-Gilbert equation is a valid approach at low temperatures only. In multiple recent applications such as heat-assisted magnetic recording or ultrafast magnetic dynamics, the temperatures often go close to the Curie temperature $T_c$ and above. Here we review the micromagnetic approach valid in this temperature range, based on the use

O. Chubykalo-Fesenko (✉)
Instituto de Ciencia de Materiales de Madrid, Spanish National Research Council – CSIC, Madrid, Spain
e-mail: oksana@icmm.csic.es

P. Nieves
International Research Center in Critical Raw Materials and Advanced Industrial Technologies, Universidad de Burgos, Burgos, Spain

of the Landau-Lifshitz-Bloch equation. The essential part of this approach is the presence of the temperature-dependent longitudinal relaxation with the characteristic time diverging at $T_c$. We review this approach in its classical and quantum formulations and for one- and two-component materials. The behavior of longitudinal relaxation time is discussed. Finally, we present examples of the use of this micromagnetics related to the modeling of ultrafast magnetization dynamics.

# 1    Introduction

Recent advances in magnetism including the design of novel nanostructured magnetic materials, better theoretical understanding of magnetic phenomena, and the development of nanoscale experimental techniques have driven the progress of nanotechnology in general but specially in the data storage industry. Recently, novel high-temperature magnetic phenomena have been discovered and attracted a lot of researchers. One of them is the laser-induced ultrafast magnetization dynamics (see, e.g., Beaurepaire et al. 1996; Vahaplar et al. 2009; Kirilyuk et al. 2010) where a subpicosecond dynamics is observed when a femtosecond laser pulse is applied to magnetic materials. Another example is the spin Seebeck effect (Uchida et al. 2008) where spin currents and spin accumulation are observed in a ferromagnet due to a temperature gradient. Apart from their fundamental interest, these phenomena are very appealing from technological perspectives that range from increasing the speed of the magnetization switching to the production of spin-voltage generators. It has been also found that a good strategy to improve the performance of hard disk devices and magnetic random-access memories is to increase the temperature of the magnetic thin film during the writing process. Therefore, it is necessary to search for models that can describe the magnetic behavior in these novel high-temperature phenomena, in which temperature is often raised up to and above the Curie temperature $T_c$.

In magnetism, micromagnetic modeling is a very useful complement to experimental measurements, especially for calculations of hysteresis and dynamics of magnetic nanostructures such as magnetic thin films, dots, stripes, etc. (Brown 1963a; Fidler and Schrefl 2000; Chantrell et al. 2001). The micromagnetics is essentially a macroscopic continuous theory. It uses a discretization of continuous magnetization function in finite elements or finite differences. The dynamics of each unit in standard micromagnetics is based on the integration of the classical Landau-Lifshitz-Gilbert (LLG) equation of motion (Landau and Lifshitz 1935; Gilbert 2004). It is essentially a zero-temperature equation, although the micromagnetic parameters could be taken as experimentally measured values at a given temperature $T$. The temperature effects are typically included by adding random fields acting on each discretization element with properties consistent with the thermodynamical equilibrium (Brown 1963b; Chubykalo et al. 2003). However, this approach is correct only for low temperatures and largely overestimates the transition temperature $T_c$ (Grinstein and Koch 2003), since the magnitude

of the magnetization vector in each element is constant. At high temperatures, high-frequency spin waves, responsible for longitudinal magnetization fluctuations near the Curie temperature $T_c$, are cut, and the value of the Curie temperature is strongly overestimated. To solve this issue, an alternative micromagnetic approach for higher temperatures based on the Landau-Lifshitz-Bloch (LLB) equation has been proposed (Garanin 1997; Chubykalo-Fesenko et al. 2006) where the magnitude of the magnetization vector is not conserved at each discretization element and the longitudinal magnetization fluctuations are introduced.

Furthermore, a complete description of the material's magnetic behavior requires very different spatial scales going from Ångström (like microscopic interactions at atomic level) to macroscale (magnetic domains typically have micrometer sizes) and also very different timescales going from femtoseconds (as in the ultrafast magnetization dynamics) to years (thermal stability in magnetic storage media). The ab initio models typically calculate microscopic parameters of the material, while atomistic spin (Heisenberg) models are good to describe the phase transitions. However, to model realistic devices such as sensors or magnetic recording head performance, large-scale modeling based on the micromagnetics is used. One way to include the effects of the microscopic properties on the magnetic macroscopic behavior is the multiscale approach (Kazantseva et al. 2008a; Atxitia et al. 2010b; Hinzke et al. 2015). Namely, ab initio calculations (the most widely used formalism is density functional theory (DFT)) are used to calculate the intrinsic parameters as magnetic moment ($\mu$), exchange constants ($J$), on-site magnetocrystalline anisotropy ($k$), etc...; then these parameters are used in atomistic classical (Heisenberg-like) models where the temperature dependence of the equilibrium magnetization $M_s(T)$, anisotropy $K(T)$, and exchange stiffness $A(T)$ among other properties can be calculated; and finally the temperature dependence of these parameters is included in the micromagnetic approach which can model the magnetic behavior at large spatial scale (see the sketch in Fig. 1). The temperature is an essential part of this approach, and the correct micromagnetic description is based on the LLB equation (Kazantseva et al. 2008a).
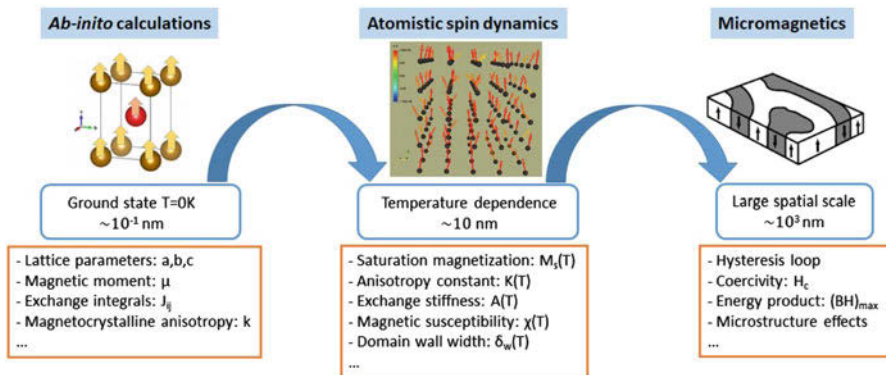


**Fig. 1** The sketch of the multiscale approach

The micromagnetic modeling based on the LLB equation has manifested itself as a good approach to model new phenomena such as the ultrafast magnetization dynamics (Kazantseva et al. 2008a; Sultan et al. 2012; Atxitia et al. 2007, 2010a, 2016; Mendil et al. 2014), spin caloritronics (Hinzke and Nowak 2011; Schlickeiser et al. 2014), and heat-assisted magnetic recording (McDaniel 2012; Greaves et al. 2015; Takano et al. 2011; Vogler et al. 2014, 2016) processes.

## 2 Classical Micromagnetic Models

### 2.1 The Classical Heisenberg (Atomistic) Modeling

The magnetic moments in the solid state can be localized or semi-localized, carried by the conduction electrons (itinerant magnetism) like in metals. However, models based on a classical Heisenberg Hamiltonian of localized spins are currently used to describe the magnetic properties in both situations. Typically, the generalized Heisenberg Hamiltonian for a uniaxial magnet in these atomistic models is written as

$$\mathcal{H} = -\sum_i \mathbf{H} \cdot \boldsymbol{\mu}_i - \sum_i k_i s_{i,z}^2 - \frac{1}{2} \sum_{i,j} J_{ij} \mathbf{s}_i \cdot \mathbf{s}_j, \tag{1}$$

where $\mathbf{H}$ is the external magnetic field, $\boldsymbol{\mu}_i$ is the magnetic moment per atom at site $i$, $\mathbf{s}_i = \boldsymbol{\mu}_i / \mu_i$, $k_i$ is the uniaxial atomic-site anisotropy constant, and $J_{ij}$ is the exchange constant between the spins $i$ and $j$. The first term in Eq. (1) corresponds to the Zeeman energy, the second term is the on-site uniaxial magnetocrystalline anisotropy energy (where the uniaxial axis is along the z-axis), and the last one is the exchange energy. The exchange energy may be long-range corresponding to the RKKY interactions and site-dependent. The dynamics of each normalized classical atomic magnetic moment $\mathbf{s}_i$ is described using the phenomenological Landau-Lifshitz-Gilbert (LLG) equation given by

$$\frac{d\mathbf{s}_i}{dt} = -\frac{\gamma}{1+\lambda^2} (\mathbf{s}_i \times \mathbf{H}_{\mathrm{eff},i}) - \frac{\gamma\lambda}{(1+\lambda^2)} \mathbf{s}_i \times (\mathbf{s}_i \times \mathbf{H}_{\mathrm{eff},i}), \tag{2}$$

where $\gamma$ is the gyromagnetic ratio, $\lambda$ is called atomic coupling to the bath (atomistic damping) parameter, $\mathbf{H}_{\mathrm{eff},i} = -\partial \mathcal{H}/\partial \boldsymbol{\mu}_i$ is the effective field, and $\mathcal{H}$ is the Hamiltonian given by Eq. (1). Equation (2) is a deterministic equation; it means that given the same initial conditions, one obtains always exactly the same dynamics. However, the atomic magnetic moment in a solid follows a stochastic dynamics due to the interaction with its surroundings. This fact is included in Eq. (2) adding a stochastic field $\boldsymbol{\zeta}_i$ to the effective field $\mathbf{H}_{\mathrm{eff},i}$, that is,

$$\frac{d\mathbf{s}_i}{dt} = -\frac{\gamma}{1+\lambda^2} (\mathbf{s}_i \times (\mathbf{H}_{\mathrm{eff},i} + \boldsymbol{\zeta}_i)) - \frac{\gamma\lambda}{(1+\lambda^2)} \mathbf{s}_i \times (\mathbf{s}_i \times (\mathbf{H}_{\mathrm{eff},i} + \boldsymbol{\zeta}_i)), \tag{3}$$

where the stochastic field has the following time average properties following the Brown's theory for nanoparticles (Brown 1963b)

$$\langle \zeta_{i,k} \rangle = 0, \quad \langle \zeta_{i,k}(0) \zeta_{i,k'}(t) \rangle = 2 \frac{\lambda k_B T}{\gamma \mu_i} \delta_{kk'} \delta(t), \quad k, k' = x, y, z, \tag{4}$$

where $k_B$ is the Boltzmann constant and $\delta$ is Kronecker delta symbol. This approach is called atomistic spin dynamics (ASD) (Skubic et al. 2008; Evans et al. 2014; Eriksson et al. 2017). The macroscopic magnetization at time $t$ is obtained as an average of the atomic magnetic moments over some volume $V$

$$\mathbf{M}(t) = \frac{1}{V} \sum_{i=1}^{N} \boldsymbol{\mu}_i(t), \tag{5}$$

where $N$ is the total number of atomic magnetic moments inside the volume $V$.

Simulations based on ASD is a powerful tool to describe the magnetic behavior of magnetic materials. Importantly, the ASD simulations lead to a good agreement with the experimentally measured Curie temperatures. Unfortunately, the size of a magnetic material that can be simulated using ASD is very limited (typically up to $(20-30 \, \text{nm})^3$ due to a large number of differential equations that must be numerically integrated.

## 2.2   Classical Micromagnetic Models at Low Temperatures

A suitable approach to study the behavior of magnetic materials at large scale is micromagnetics. It is based on the continuum approximation where the length scales considered are large enough for the atomic structure of the material to be ignored and small enough to resolve magnetic structures such as domain walls or vortices. In the continuum approximation the macroscopic magnetization is assumed to be a spatial continuous function over the material

$$\mathbf{M}(\mathbf{r}) = M_s \mathbf{m}(\mathbf{r}), \tag{6}$$

where $|\mathbf{m}(\mathbf{r})| = 1$ and $M_s$ is the saturation magnetization. In this approach the energy of a uniaxial magnetic system reads (e.g., Coey 2009)

$$E = \int_V d\mathbf{r} \{ -\mathbf{M} \cdot \mathbf{H} - K m_z^2 + A(\nabla \mathbf{m})^2 - \frac{1}{2} \mathbf{M} \cdot \mathbf{H}_d \}, \tag{7}$$

where $K$ is the macroscopic uniaxial anisotropy constant, $A$ is the exchange stiffness parameter, $(\nabla \mathbf{m})^2 = (\nabla m_x)^2 + (\nabla m_y)^2 + (\nabla m_z)^2$, $\nabla$ is the gradient operator, and $\mathbf{H}_d$ is the magnetostatic field which must be calculated self-consistently.
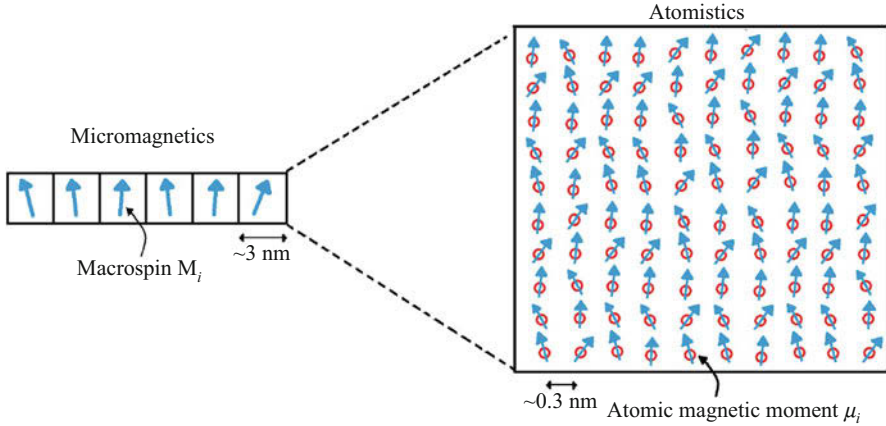
**Fig. 2** Illustration of the relation between atomistic model and micromagnetics. (Reprinted from Nieves 2015)

In order to solve micromagnetic problems numerically, the system is divided in cells of volume $V_i$ (typically around few nm$^3$), and then the average magnetization of the cell at position $\mathbf{r}_i$ is represented by vector called macrospin (see Fig. 2), $\mathbf{M}(\mathbf{r}_i) = \mathbf{M}_i = M_s \mathbf{m}_i$. In a coarse-grained (multiscale) approach, the mascrospin is the average of atomic spin moments in the sense of Eq. (5). In order to choose a suitable volume cell, it is important to take into account the domain wall width parameter $\delta$ and the exchange length $l_{\text{ex}}$ which are given by

$$\delta = \pi \sqrt{\frac{A}{K}} \ , \qquad l_{\text{ex}} = \sqrt{\frac{A}{M_s^2}}. \tag{8}$$

The domain wall width parameter corresponds to the width of a Bloch wall that is found in magnetic materials with a large magnetocrystalline anisotropy. The exchange length is the length below which atomic exchange interactions dominate typical magnetostatic fields, and it is proportional to the Néel domain wall width. The discretization length must be less than the domain wall width but includes enough atoms to be valid as the continuous approximation. The main difficulty of micromagnetics is the correct calculation of the long-range magnetostatic fields.

The standard dynamic micromagnetics is based on the same LLG equation as in the atomistic approach (however with more justified damping form) for the macrospin $\mathbf{M}_i$ given by

$$\frac{d\mathbf{M}_i}{dt} = -\frac{\gamma}{1 + \alpha_{\text{LLG}}^2}(\mathbf{M}_i \times \mathbf{H}_{\text{eff},i}) - \frac{\gamma \alpha_{\text{LLG}}}{(1 + \alpha_{\text{LLG}}^2)M_s}\mathbf{M}_i \times (\mathbf{M}_i \times \mathbf{H}_{\text{eff},i}), \tag{9}$$

where $\alpha_{\text{LLG}}$ is called the LLG damping (not to be confused with the microscopic damping in the atomistic approach, often designated with the same letter $\alpha$ but called

here $\lambda$ in order to stress the difference) and $\mathbf{H}_{\text{eff},i}$ is the effective field which is given by

$$\mathbf{H}_{\text{eff},i} = -\frac{1}{V_i}\frac{\partial E}{\partial \mathbf{M}_i} = \mathbf{H}_i + \frac{2K_i m_{i,z}}{M_s}\mathbf{e}_z + \frac{2A_i}{M_s}\nabla^2 \mathbf{m}_i + \mathbf{H}_{d,i}. \tag{10}$$

The temperature effects in micromagnetism are typically included in the following way:

(i) *Via temperature-dependent parameters:* The parameters $M_s(T)$, $A(T)$, $K(T)$, $\alpha_{\text{LLG}}(T)$ are taken at given temperature. Importantly the damping parameter $\alpha_{\text{LLG}}$ is also temperature-dependent which is frequently forgotten. As it was mentioned in the introduction, their temperature dependence can be calculated numerically using ASD within the multiscale approach; a detailed explanation of this calculation can be found, for example, in Kazantseva (2008) for FePt. They can also be obtained theoretically, for example, using the mean-field approximation (MFA) or measured experimentally.

(ii) *Via thermal field:* In 1963 W.F. Brown (1963b), considering the superparamagnetism problem in a collection of non-interacting nanoparticles, suggested to include thermal fluctuations in the LLG dynamical equation as stochastic fields whose properties are defined by the equilibrium solution of the corresponding Fokker-Planck (FP) equation. Importantly, these fields are just formal quantities and do not have physical sense; they are used in order to produce stochastic deviations of different magnetization trajectories from their averaged value with correct statistical properties. In 1993 Lyberatos and Chantrell (1993) for the first time studied the dynamics of two interacting magnetic dipoles including these fluctuating thermal fields. This idea was further developed by many authors (Nakatani et al. 1997; García-Palacios and Lázaro 1998; Scholz et al. 2001; Chubykalo et al. 2002). It was shown that the stochastic fields remain uncorrelated and with the same properties for the interacting case (Chubykalo et al. 2003). These developments lead to Langevin dynamics micromagnetics where a fluctuating thermal field $\boldsymbol{\zeta}_i$ is added to the effective field given by Eq. (10) with the following time average properties (the same as used in the ASD dynamics)

$$\langle \zeta_{i,k}\rangle_t = 0, \quad \langle \zeta_{i,k}(0)\zeta_{i,k'}(t)\rangle_t = 2\frac{\alpha_{\text{LLG}}k_B T}{\gamma M_s V_i}\delta_{kk'}\delta(t), \quad k, k' = x, y, z. \tag{11}$$

The main feature of this approach is that the magnitude of every macrospin is conserved in all dynamical processes, that is, $|\mathbf{M}_i| = $const. However, the direct comparison with the ASD shows that thermal simulations of the magnetization dynamics based on the micromagnetic LLG equation are not suitable for high temperatures (Chubykalo-Fesenko et al. 2006). This is due to the fact that micromagnetic simulations do not include the high-frequency spin waves, and, thus, the Curie temperature is seriously overestimated (Grinstein and Koch 2003). Additionally, in recent ASD simulations (Chubykalo-Fesenko et al. 2006; Kazantseva et al. 2007),

it has been demonstrated that at high temperatures several important effects occur which cannot be taken into account in the micromagnetic LLG approach. Namely, during the magnetization dynamics, (i) the magnetization vector magnitude is not conserved, (ii) longitudinal magnetization relaxation occurs with the longitudinal relaxation time increment approaching $T_C$ (known as critical slowing down), and (iii) at the same time the transverse relaxation time decreases (Chubykalo-Fesenko et al. 2006). Therefore, a different micromagnetic approach is required at elevated temperatures.

## 2.3    Thermal Micromagnetics at High Temperatures

In 1997 an alternative approach was suggested by D. Garanin (1997). Based on the Fokker-Planck equation, he derived a classical macroscopic equation of motion for the magnetization called Landau-Lifshitz-Bloch (LLB) equation. The name was chosen in order to stress that the magnetization behavior interpolates between the Landau-Lifshitz equation at low temperatures (micromagnetic LLG Eq. (9)) and the well-known Bloch equation (Bloch 1946) at high temperatures. The Bloch equation is a phenomenological equation frequently used in the nuclear spin resonance description, for example, for protons in a water molecule. The dynamical equation for the ensemble of paramagnetic spins involves the same precessional term and two phenomenological relaxational parameters describing the longitudinal relaxation (defined by the characteristic time $T_1$) and the transverse relaxation ($T_2$). In the nuclear spin resonance case $T_2 \ll T_1$, for ferromagnetic materials of the interest here, the situation is the opposite due to a dominant role of large exchange interactions.

The derivation of the classical LLB assumes the classical atomistic approach (ASD) of Sect. 2.1, only based on the Landau-Lifshitz (LL) equation instead of the LLG (i.e., disregarding the term $\lambda^2$ in Eq. (2), which anyway is small). First, the Fokker-Planck equation corresponding to many-spin Eq. (9) was calculated in Garanin (1997) (see also details in Atxitia 2012). Using the dynamics of the probability function, governed by this equation and disregarding the third-order moment distribution, one obtains an equation for thermal average of the spin polarization, i.e., the reduced magnetization $\mathbf{m} = < \mathbf{s}_i >$ in a paramagnetic state. For the treatment of ferromagnet, the external field is substituted by the mean field.

The corresponding LLB equation for $\mathbf{m}$ has the following form (see details in Garanin 1997 and Atxitia 2012):

$$\frac{d\mathbf{m}}{dt} = -\gamma[\mathbf{m} \times \mathbf{H}^{\text{MFA}}] - \Gamma_\parallel \left(1 - \frac{\mathbf{m}\mathbf{m}_0}{m^2}\right)\mathbf{m} - \Gamma_\perp \frac{[\mathbf{m} \times [\mathbf{m} \times \mathbf{m}_0]]}{m^2}, \qquad (12)$$

where

$$\mathbf{m}_0 = L(\xi)\frac{\boldsymbol{\xi}}{\xi}, \quad \boldsymbol{\xi} \equiv \beta\mu\mathbf{H}^{\text{MFA}}, \quad \mathbf{H}^{\text{MFA}} = \frac{zJ}{\mu}\mathbf{m} + \mathbf{H}'. \qquad (13)$$

Here $\xi \equiv |\boldsymbol{\xi}|$, $L(\xi) = \coth(\xi) - 1/\xi$ is the Langevin function, $\beta = 1/k_B T$, $k_B$ is the Boltzmann constant; $\mu$ is the atomic magnetic moment; $z$ is the number of nearest neighbors; $J$ is the Heisenberg exchange interaction parameter; $\mathbf{H}'$ contains the external magnetic field and the nonhomogeneous part of the exchange interaction;

$$\Gamma_{\parallel} = \Lambda_N \frac{L(\xi)}{\xi L'(\xi)}, \quad \Gamma_{\perp} = \frac{\Lambda_N}{2} \left( \frac{\xi}{L(\xi)} - 1 \right) \tag{14}$$

describe parallel and perpendicular relaxation rates, respectively; $\Lambda_N = 2\gamma \lambda k_B T/\mu$ is the characteristic diffusion relaxation rate or, for the thermo-activation escape problem, the Néel attempt frequency; and $L'(\xi) = dL/d\xi$ is the derivative of the Langevin function. This equation can be already used for modeling and in many cases gives even a better agreement with the ASD than a more conventional LLB equation presented below.

To put this equation in the form, similar to the LLG one, the homogeneous part of the exchange field is assumed to be much larger than all other fields acting in the system ($[zJ/\mu]m \gg H'$). This leads to the conventional form of the LLB equation which in micromagnetics is written for each discretization element (macrospin) $i$

$$\frac{d\mathbf{m}_i}{dt} = -\gamma \left[ \mathbf{m}_i \times \mathbf{H}_{\text{eff}}^i \right] + \frac{\gamma \alpha_{\parallel}}{m_i^2} \left( \mathbf{m}_i \cdot \mathbf{H}_{\text{eff}}^i \right) \mathbf{m}_i - \frac{\gamma \alpha_{\perp}}{m_i^2} \left[ \mathbf{m}_i \times \left[ \mathbf{m}_i \times \mathbf{H}_{\text{eff}}^i \right] \right], \tag{15}$$

where $\mathbf{m}_i = \mathbf{M}_i/M_e(0)$ with $M_e(0) = M_s(0)$ being the equilibrium saturation magnetization at $T = 0K$. The longitudinal and transverse relaxation parameters are

$$\alpha_{\parallel} = \lambda \frac{2T}{3T_c}, \qquad \alpha_{\perp} = \lambda \cdot \begin{cases} \left[ 1 - \frac{T}{3T_c} \right] & T \lesssim T_c, \\ \frac{2T}{3T_c} & T \gtrsim T_c. \end{cases} \tag{16}$$

Note that in the alternative representation, the magnetization can be normalized at its value at $T$, i.e., using $\mathbf{n}_i = \mathbf{M}_i/M_e(T)$ as is typically done for the LLG micromagnetics. This leads to the same form of the LLB equation but now for the variable $\mathbf{n}_i$. The only difference is the renormalization of the damping parameters by the factor $m_e = M_e(T)/M_e(0)$. Since $m_e \to 0$ at $T_c$, this form of the LLB equation is obviously not useful for modeling near the phase transition but should be used for the comparison with the LLG micromagnetic modeling.

In the LLB-based micromagnetic approach, the effective fields are given by

$$\mathbf{H}_{\text{eff}}^i = \mathbf{H} + \mathbf{H}_{i,\text{EX}} + \mathbf{H}_{i,A} + \begin{cases} \frac{1}{2\tilde{\chi}_{\parallel}} \left( 1 - \frac{m_i^2}{m_e^2} \right) \mathbf{m}_i & T \lesssim T_c, \\ \frac{J_0}{\mu} \left( 1 - \frac{T}{T_c} - \frac{3m_i^2}{5} \right) \mathbf{m}_i & T \gtrsim T_c, \end{cases} \tag{17}$$

where $\widetilde{\chi}_{\parallel} = (\partial m/\partial H)_{H\to\infty}$ is the reduced longitudinal susceptibility and $J_0$ is the zero Fourier component of the exchange interaction which is related to the Curie temperature $T_c$ in the MFA through $T_c = J_0/(3k_B)$ (in the simple cubic lattice case with nearest interactions only $J_0 = 6J$). The last term in Eq. (17) describes the internal exchange field inside the macrospin. $\mathbf{H}$ is the applied magnetic field and $\mathbf{H}_{i,\mathrm{EX}}$ is exchange interaction between macrospins, and it is given by

$$\mathbf{H}_{i,\mathrm{EX}} = \frac{2A_i(T)}{m_i^2 M_s(0)}\nabla^2\mathbf{m}_i = \frac{2A_i(T)}{m_i^2 M_s(0)\triangle^2}\sum_{\langle i,j\rangle}\left(\mathbf{m}_j - \mathbf{m}_i\right), \tag{18}$$

where $\langle i, j\rangle$ means a sum over neighbors, $\triangle$ is the lateral size of the micromagnetic discretization cell, $A_i(T)$ is the micromagnetic exchange also called stiffness parameter, and $\mathbf{H}_{i,A}$ is the anisotropy field given by

$$\mathbf{H}_{i,A} = -\frac{1}{\widetilde{\chi}_{\perp}}\left(m_{i,x}\mathbf{e}_x + m_{i,y}\mathbf{e}_y\right), \tag{19}$$

where $\widetilde{\chi}_{\perp}$ is the reduced transverse susceptibility. Note that the use of the transverse susceptibility is more physically correct as that of the anisotropy (see discussion in Garanin and Chubykalo-Fesenko 2004). However, in practice the anisotropy notion is still used in most of cases, i.e., one assumes $\widetilde{\chi}_{\perp}(T) = M^2/[2K(T)]$.

The LLB equation contains six temperature-dependent parameters: $\lambda(T)$, $M_e(T)$, $A(T)$, $K(T)$, $\widetilde{\chi}_{\parallel}(T)$, and $\widetilde{\chi}_{\perp}(T)$. Typically, the parameter $\lambda$ is considered in ASD temperature independent. This however is not true in the ab initio modeling (see also discussion in Nieves et al. 2014). In principle other parameters can be measured experimentally. In the multiscale approach the temperature dependence of $M_e(T)$, $A(T)$ and $K(T)$ is calculated using ASD (Kazantseva et al. 2008a; Atxitia et al. 2010b; Nieves et al. 2017), and the scaling with magnetization relations can be used for the last two of them (Kazantseva 2008; Atxitia et al. 2010b; Moreno et al. 2016; Asselin et al. 2010). The parameter $M_e(T)$ can be also obtained using the MFA as the solution of the equation $m_e = L(\beta J_0 m_e)$ where $L(x) = \coth(x) - 1/x$ is the Langevin function. The exchange parameter $J_0$ can be calculated by first principles or roughly estimated by MFA expression $J_0 = 3k_B T_c$ using experimental Curie temperature. The reduced longitudinal susceptibility can be calculated either by using MFA as

$$\widetilde{\chi}_{\parallel}(T) = \begin{cases} \dfrac{\mu_{at}\beta L'}{1 - \beta J_0 L'} & T \lesssim T_c, \\[2mm] \dfrac{\mu_{at}T_c}{J_0(T - T_c)} & T \gtrsim T_c, \end{cases} \tag{20}$$

where $L'$ is the derivative of the Langevin function evaluated at $\beta J_0 m_e$, or by ASD simulations (Hinzke et al. 2000).

The first term in Eq. (15) describes the precession of magnetization vector $\mathbf{m}$ around its effective field $\mathbf{H}_{\mathrm{eff}}$ (see Fig. 3a), and the second and the third terms
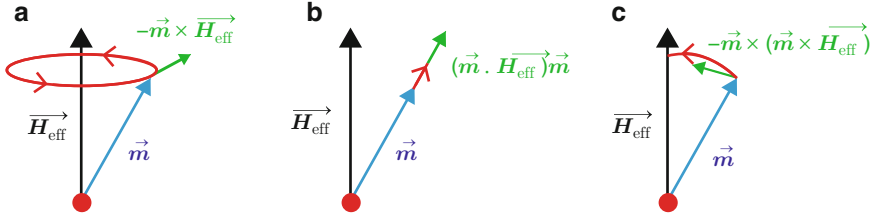
**Fig. 3** Diagram illustrating the meaning of three terms in the LLB equation: (**a**) precession, (**b**) longitudinal dynamics, and (**c**) transverse dynamics. (Reprinted from Nieves 2015)

describe the longitudinal (see Fig. 3b) and the transverse (see Fig. 3c) dynamics, respectively. Comparing the LLB equation (15) with LLG equation (9), we notice that the former contains an extra term which describes the longitudinal relaxation, that is, it drives the dynamics of the magnitude of **m** giving a more accurate description of the magnetic dynamics at elevated temperatures. In particular this extra term plays a crucial role in ultrafast magnetization dynamics, i.e., at the timescale below 1 ps. At low temperatures the additional internal field in Eq. (17) is very large (since the susceptibility is very small) and keeps the magnetization length constant. Since the LLB equation contains the LLG one, it also describes the damped precession which typically takes place between 0.1 and 10 ns (see Fig. 3a).

The effective field of the LLB equation given by Eq. (17) can be also obtained as

$$\mathbf{H}_{\text{eff}}^i = -\frac{1}{V_i} \frac{\partial \mathcal{F}}{\partial \mathbf{M}_i}, \tag{21}$$

where $\mathcal{F}$ is the free energy given by

$$\mathcal{F}(\mathbf{M}, T) = \mathcal{F}_0 + \int_V d\mathbf{r} \{-\mathbf{M} \cdot \mathbf{H} + \frac{1}{2\chi_\perp}(M_x^2 + M_y^2) + A \left( \nabla \left[ \frac{\mathbf{M}}{M} \right] \right)^2$$

$$+ \frac{1}{8 M_e^2(T)\chi_\parallel}[M^2 - M_e^2(T)]^2\}, \tag{22}$$

where $\mathcal{F}_0$ is the equilibrium free energy in the absence of anisotropy and magnetic field, $\chi_\parallel = M_s(0)\widetilde{\chi}_\parallel$ and $\chi_\perp = M_s(0)\widetilde{\chi}_\perp$. In Kachkachi and Garanin (2001) derived this free energy in the whole temperature range using a procedure based on the MFA. Obviously close to the transition temperature, this functional is reduced to the well-known Landau-Ginzburg expression. This derivation further justifies the thermodynamic consistency of the LLB equation.

Similarly as it was done in the micromagnetic LLG equation, Garanin and Chubykalo-Fesenko (2004) included stochastic thermal fields into the LLB equation. However, Evans et al. (2012) noticed that this approach doesn't recover the Boltzmann distribution close to $T_c$ at equilibrium. In order to solve this issue, they suggested an alternative stochastic LLB equation of the form

$$\frac{d\mathbf{m}_i}{dt} = -\gamma \left[ \mathbf{m}_i \times \mathbf{H}_{\text{eff}}^i \right] + \frac{\gamma \alpha_\parallel}{m_i^2} \left( \mathbf{m}_i \cdot \mathbf{H}_{\text{eff}}^i \right) \mathbf{m}_i$$

$$- \frac{\gamma \alpha_\perp}{m_i^2} \left[ \mathbf{m}_i \times \left[ \mathbf{m}_i \times \left( \mathbf{H}_{\text{eff}}^i + \boldsymbol{\zeta}_{i,\perp} \right) \right] \right] + \boldsymbol{\zeta}_{i,ad}, \qquad (23)$$

with two thermal fields: a multiplicative transverse $\boldsymbol{\zeta}_{i,\perp}$ and an additive longitudinal $\boldsymbol{\zeta}_{i,ad}$ given by

$$\left\langle \zeta_{i,\perp}^k(0) \zeta_{j,\perp}^l(t) \right\rangle = \frac{2\gamma k_B T \left( \alpha_\perp - \alpha_\parallel \right)}{M_e(0) V_i \alpha_\perp^2} \delta_{ij} \delta_{kl} \delta(t) \qquad (24)$$

$$\left\langle \zeta_{i,ad}^k(0) \zeta_{j,ad}^l(t) \right\rangle = \frac{2\gamma k_B T \alpha_\parallel}{M_e(0) V_i} \delta_{ij} \delta_{kl} \delta(t), \qquad (25)$$

where $i$ and $j$ denote macrospin index and $k$ and $l$ denote the Cartesian components x, y, and z. Thus in order to describe the thermal dispersion of the magnetization trajectories, we now have a Langevin dynamics simulations based on the LLB equation. The dispersion of magnetization trajectories is very important for modeling of magnetization dynamics under the action of thermal laser pulse (Kazantseva et al. 2008a). Note that not very close to $T_c$ the approaches of Garanin and Chubykalo-Fesenko (2004) and Evans et al. (2012) are indistinguishable.

The comparison of the macrospin LLB equation with direct ASD simulations gives a very good agreement (Chubykalo-Fesenko et al. 2006; Kazantseva et al. 2008a). Note that a typical problem of the thermal micromagnetics is the presence of finite-size effects, i.e., the dependence of the average magnetization on the discretization size. Comparatively to the LLG approach, thermal micromagnetics based on the LLB equation have these effects largely reduced (Atxitia et al. 2007). This happens due to the fact that the LLB equation forces the equilibrium solution to be $M_e$. However, the finite-size effects are an inherent part of the thermal micromagnetic approach and are always present (Atxitia et al. 2007). If these effects are taken into account, the agreement between the atomistic and LLB-based micromagnetics can be even better (Vogler et al. 2014, 2016).

## 3 The Quantum (Semiclassical) Landau-Lifshitz-Bloch Equation

The above LLB equation is based on the ASD approach. There is also a quantum derivation, which was published even earlier than the classical one by Garanin et al. (1990) and Garanin (1991). Obviously, the quantum derivation can be made for simplified models only. In the original derivation, this was done for an isolated spin interacting with a phonon bath via direct and Raman processes with the Hamiltonian presented in Fig. 4. Lately, this was also done for a spin interacting with a simple electron bath by Nieves et al. (2014) following a simple model of Koopmans et al. (2005) conceived for ultrafast magnetization dynamics.

Hamiltonian:

Scattering probabilities

$$\hat{H} = -\mathrm{H}(t)\cdot\hat{\mathrm{S}} + \sum_q w_q \hat{a}_q^+ \hat{a}_q - \sum_q V_q(\eta\cdot\hat{\mathrm{S}})(\hat{a}_q^+ + \hat{a}_{-q}) - \sum_{p,q} V_{p,q}(\eta\cdot\hat{\mathrm{S}})\hat{a}_p^+ \hat{a}_q$$

Zeeman    Phonon bath    Direct transformation    Raman processes

$|\sigma'\rangle$ $\quad$ $|q\rangle$ $|\sigma'\rangle$ $\qquad$ $|\sigma'\rangle$ $\quad V_{p,q}$ $\quad |p\rangle$

$V_q$ $\qquad$ $V_q$

$\qquad + \qquad$

$|\sigma\rangle \qquad |\sigma\rangle \quad |-q\rangle \qquad |\sigma\rangle \qquad |q\rangle$
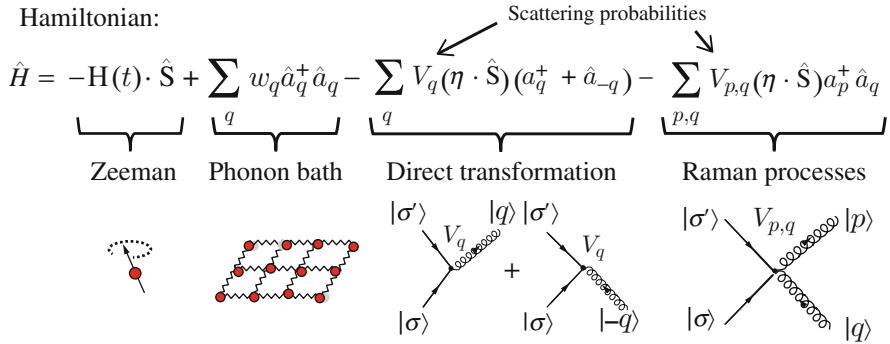
**Fig. 4** The sketch of the quantum model based on the spin-phonon scattering. (Reprinted from Nieves 2015)

The derivation of the quantum LLB (qLLB) equation by Garanin (1991) is based on a standard density matrix approach (see, e.g., Blum 1981; Garanin 2012) for a system interacting weakly with a bath. The sketch for the spin-phonon model is presented in Fig. 4, and the spin-electron model can be found in Nieves et al. (2014). Namely, starting from the Schrödinger equation, one can obtain a Liouville equation for the time evolution of the density operator $\hat{\rho} = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle$ is the wave function of the whole system (spin and phonons or spin and electrons). Next, the bath is assumed to be in a quasi-equilibrium meaning that at this timescale the phonon (electron) distribution can be described by the Bose-Einstein (Fermi-Dirac) one with a defined temperature (Nieves et al. 2014) although the temperature is still slow varying in time. The temperature dynamics can be external, for example, described by the two-temperature model (Atxitia et al. 2007; Nieves and Chubykalo-Fesenko 2016; John et al. 2017; Mendil et al. 2014). Note that in the self-consistent description, the magnetization dynamics also produces a temperature dynamics and vice versa (Nieves et al. 2016).

Furthermore, the interactions with the bath are assumed to be small so that they cannot cause a significant entanglement between both systems; this allows to factorize the density operator by its spin and bath part $\hat{\rho}(t) \cong \hat{\rho}_s(t)\hat{\rho}_b^{eq}$ and average over the bath variable (Garanin 2012). The following approximations are also made: (i) the Markov or short memory approximation assuming that the spin dynamics is slower than the phonon (electron) dynamics; (ii) the secular approximation, which consists in neglecting the fast oscillating terms; (iii) as in the classical case, the interactions are taken into account in the MFA, the homogeneous part of the exchange field is considered to be much larger than all other approximations, and close to $T_c$ the expansion $H_\mathrm{E} \simeq (J_0/\mu_\mathrm{at})(m_e + \widetilde{\chi}_\parallel h)$ is used.

The final formal form of the qLLB equation is the same as in the classical case (15), the equilibrium magnetization is defined now by the Brillouin function for the spin S: $m_e = B_S(\beta J_0 m_e)$, and the MFA longitudinal susceptibility $\widetilde{\chi}_\parallel$ follows from it at $T < T_c$ as $\widetilde{\chi}_\parallel = \mu_{at}\beta B'_S/(1 - \beta B'_S J_0)$ where $B'_S(x) = dB_S/dx$ is

evaluated at the equilibrium $B'_S = B'_S(\beta J_0 m_e)$. The parameters $\alpha_\parallel$ and $\alpha_\perp$ are different and can be conveniently expressed in a form which is suitable for the comparison with the classical LLB equation. Below $T_c$ the damping parameters are written as

$$\alpha_\parallel = \lambda \frac{2T}{3T_c} \frac{2q_s}{\sinh(2q_s)} \tag{26}$$

$$\alpha_\perp = \lambda \left[ \frac{\tanh(q_s)}{q_s} - \frac{2T}{3T_c} \left( 1 - \frac{K_1}{2K_2} \right) \right], \tag{27}$$

where $q_s = 3T_c m_e / (2(S+1)T)$ and

$$\lambda = K_2 \frac{(S+1)}{S} \frac{\mu_{at}}{\gamma k_B T}. \tag{28}$$

has the sense of atomistic damping (coupling to the bath) parameter. The parameters $K_1$ and $K_2$ are related to the microscopic scattering probabilities (see Nieves et al. 2014).

Below $T_c$ the effective field used in Eq. (15) is given by Eq. (17). Above $T_c$ we also rewrite the effective field in terms of the longitudinal susceptibility at $T > T_c$, i.e., $\tilde{\chi}_\parallel = \mu T_c / [J_0(T - T_c)]$, and the field reads as

$$\mathbf{H}_{\text{eff}} = -\frac{1}{\tilde{\chi}_\parallel} \left( 1 + \frac{3T_c m^2}{5A_s(T - T_c)} \right) \mathbf{m} + \mathbf{h} \ , \ \frac{T_c}{T - T_c} \gg 1, \tag{29}$$

where $A_s = 2(S+1)^2 / ([S+1]^2 + S^2)$ and $\mathbf{h}$ contains all other fields (Zeeman, anisotropy, (external) inter-macrospin exchange, and magnetostatic). Note that although $\tilde{\chi}_\parallel$ is divergent at $T_c$ as it corresponds to the second-order phase transition, the internal exchange field is the same for any $T_c - \varepsilon$ and $T_c + \varepsilon$ insuring that under the integration of the LLB equation, $\mathbf{m}(t)$ rests continuous through the critical point. In the region just above $T_c$, $q_s = 0$ and $K_1 \cong K_2$, so that the damping parameters become approximately the same and equal to the one presented in Eq. (16), where the dependence on the spin value $S$ is included implicitly through $\lambda$ (see Eq. (28)).

In the special case with $S = 1/2$ and pure longitudinal dynamics, the qLLB equation is reduced (see details in Nieves et al. 2014) to the so-called self-consistent Bloch equation (see Xu and Zhang 2012)

$$\frac{d\mathbf{m}}{dt} = -\gamma \mathbf{m} \times \mathbf{h} - \frac{\mathbf{m} - \mathbf{m}_0}{\tau_s}, \tag{30}$$

where $\mathbf{m}_0 = B_{1/2}(\xi)\boldsymbol{\xi}/\xi$ and $\tau_s = 1/K_2$ the spin relaxation time. This equation may be further simplified taking into account that the exchange field is large in which case it is reduced to

$$\frac{dm}{dt} = \frac{m}{\tau_s}\left[1 - m\coth\left(\frac{mT_c}{T}\right)\right].$$  (31)

The above equation was used as a part of the so-called three-temperature model (M3TM) by Koopmans et al. (2010), in which $\tau_s$ is related to the Elliott-Yafet scattering probability.

## 4    The Two-Sublattice Landau-Lifshitz-Bloch Equation

For the treatment of a two-sublattice ferro- or ferrimagnet by Atxitia et al. (2012), two Eq. (12) are written for each sublattice separately, and the mean field contains inter- and intra-sublattice contributions. For the disordered case with different concentrations of the species (such as the case of $Gd_xCo_{1-x}$, see Fig. 5), the MFA field reads:

$$\mathbf{H}_v^{MFA} = \frac{J_{0,v}}{\mu_v}\mathbf{m}_v + \frac{J_{0,v\kappa}}{\mu_v}\mathbf{m}_\kappa + \mathbf{h}_v,$$  (32)

where $J_{0,v} = x_v z J_{vv}$, $J_{0,v\kappa} = x_\kappa z J_{v\kappa}$; $z$ is the number of nearest neighbors in the ordered lattice; $J_{vv}$ and $J_{v\kappa}$ are the Heisenberg intra- and intersublattice exchange interaction parameters; $x_v$ and $x_\kappa = 1 - x_v$ are the concentrations of the sublattices $v$ and $\kappa$, respectively; and $\mathbf{h}_v$ contains the external applied and the anisotropy fields acting on the sublattice $v$. With the same approximations as in the case of the classical one sublattice LLB equation, one arrives at two coupled LLB equations of the form Eq. (15) for each sublattice, where the effective field is given by

$$\mathbf{H}_{eff,v} = \mathbf{H} + \mathbf{H}_{A,v} + \frac{J_{0,v\kappa}}{\mu_v}\mathbf{\Pi}_\kappa$$  (33)

$$- \left[\frac{1}{\Lambda_{vv}}\left(m_v - m_{e,v}\right) - \frac{1}{\Lambda_{v\kappa}}\left(|\tau_\kappa| - |\tau_{e,\kappa}|\right)\right]\frac{\mathbf{m}_v}{m_v},$$



**Fig. 5** Sketch of a disordered magnetic alloy from the point of view of the atomistic model (left panel) and the micromagnetic LLB model (right panel). (Reprinted from Atxitia et al. 2012)
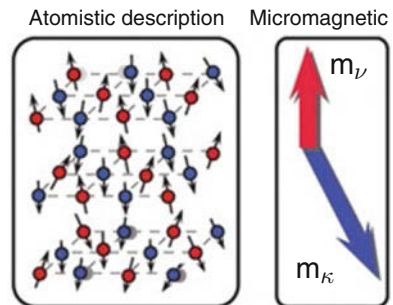
where $\mathbf{H}_{A,v}$ is the anisotropy field, $\mathbf{H}$ is the applied field, $\mathbf{\Pi}_\kappa = -[\mathbf{m}_v \times [\mathbf{m}_v \times \mathbf{m}_\kappa]]/m_v^2$, $\tau_v = (\mathbf{m}_v \cdot \mathbf{m}_\kappa)/m_\kappa$, $\tau_{e,v} = (\mathbf{m}_{e,v} \cdot \mathbf{m}_{e,\kappa})/m_{e,\kappa}$, and

$$\Lambda_{vv}^{-1} = \frac{1}{\widetilde{\chi}_{v,||}} \left( 1 + \frac{J_{0,v\kappa}}{\mu_v} \widetilde{\chi}_{\kappa,||} \right), \quad \Lambda_{v\kappa}^{-1} = \frac{|J_{0,v\kappa}|}{\mu_v}. \tag{34}$$

The longitudinal susceptibility can be calculated in the MFA as in Atxitia et al. (2012)

$$\widetilde{\chi}_{v,||} = \frac{\mu_\kappa \beta L'_v J_{0,v\kappa} \beta L'_\kappa + \mu_v \beta L'_v \left( 1 - J_{0,\kappa} \beta L'_\kappa \right)}{\left( 1 - J_{0,v} \beta L'_v \right) \left( 1 - J_{0,\kappa} \beta L'_\kappa \right) - J_{0,\kappa v} \beta L'_v J_{0,v\kappa} \beta L'_\kappa}, \tag{35}$$

where $L'_v = L'(\xi_{e,v})$ with $\xi_{e,v} = \beta(J_{0,v} m_{e,v} + |J_{0,v\kappa}| m_{e,\kappa})$. The damping parameters are

$$\alpha_{||}^v = \frac{2\lambda_v}{\beta \widetilde{J}_{0,v,e}}, \quad \alpha_\perp^v = \lambda_v \left( 1 - \frac{1}{\beta \widetilde{J}_{0,v,e}} \right), \tag{36}$$

where $\widetilde{J}_{0,v,e} = J_{0,v} + |J_{0,v\kappa}| (m_{e,\kappa}/m_{e,v})$. Note that all the expressions above are the same for two-sublattice ferrimagnetic and ferromagnetic alloys.

The quantum case only differs by the damping parameters which now have the forms (see Nieves 2015):

$$\alpha_{||}^v = \frac{2\lambda_v}{\beta \tilde{J}_{0,v,e}} \left( \frac{S_v}{S_v + 1} \right) \frac{2q_v}{\sinh(2q_v)}, \tag{37}$$

$$\alpha_\perp^v = \lambda_v \left[ \frac{\tanh(q_v)}{q_v} - \frac{2S_v}{(S_v + 1)\beta \tilde{J}_{0,v,e}} \left( 1 - \frac{K_{1,v}}{2K_{2,v}} \right) \right], \tag{38}$$

where $q_v = (\beta \tilde{J}_{0,v,e} m_{e,v})/(2S_v)$ and $\lambda_v = [\beta \mu_v K_{2,v}(S_v + 1)]/[\gamma_v S_v]$.

In the equations above the longitudinal susceptibility $\widetilde{\chi}_{v,||}$ should be evaluated in MFA using the Brillouin function.

Taking the limits $S_v \to \infty$ and $S_\kappa \to \infty$ in the quantum version, we arrive to the classical LLB equation for disordered magnetic alloys. On the other hand, if we take the limit $x_\kappa \to 0$ (i.e., the impurities are removed), then we obtain the qLLB equation for a ferromagnet. Very recently, Vogler et al. (2018) incorporated stochastic fluctuations to the classical two-sublattice LLB.

## 4.1    The Relaxation Rates

The main difference of the high-temperature micromagnetics based on the LLB equation and the conventional micromagnetics is the temperature-dependent relaxation. It is defined by the two main characteristic times, obtained by the linearization. These are the longitudinal relaxation time

$$\tau_{\parallel} = \frac{\widetilde{\chi}_{\parallel}}{\gamma \alpha_{\parallel}}, \tag{39}$$

which defines the speed of the change of the magnetization length and the transverse relaxation time $\tau_{\perp}$, i.e., the characteristic time taken by the transverse component of magnetization to relax to the effective field **h** including the external field and the anisotropy contributions

$$\tau_{\perp} = \frac{\gamma \, h \, m_e}{\alpha_{\perp}}. \tag{40}$$

The corresponding transverse relaxation below $T_c$ may be put in the more common form of the macroscopic LLG equation. For this instead of the normalization of magnetization to the total spin polarization, one should use its normalization to the saturation magnetization value, i.e., $M_e(T)$. The resulting equation is the same LLB one but with a different damping parameters, called here $\alpha_{\text{LLG}}$.

$$\alpha_{\text{LLG}} = \frac{\alpha_{\perp}}{m_e}. \tag{41}$$

Note that while both $\alpha_{\parallel}$ and $\alpha_{\perp}$ are continuous through $T_c$, $\tau_{\parallel}$ and $\alpha_{LLG}$ diverge at $T_c$, corresponding to the critical behavior at the phase transition. At the same time the perpendicular relaxation time goes to zero which constitutes one of the main differences between the LLG and the LLB dynamics.

The most important manifestation of the high-temperature dynamics is the longitudinal relaxation time. It will show itself only close to the phase transition due to the fact that at relatively low temperatures the longitudinal susceptibility is small. At $T_c$ the susceptibility diverges as well as the longitudinal relaxation time. This divergence is only visible very close to $T_c$ so that for practical reasons the longitudinal timescale is still at femtosecond scale for $T/T_c = 0.9$. However, this timescale is now accessible with the ultrafast magnetization dynamics exited by fs laser or THz radiation.

An example of the temperature dependence of the longitudinal relaxation time is presented in Fig. 6. It has the following asymptotic behavior:

$$\tau_{||} \cong \frac{\mu_{at}}{2\gamma \lambda k_B T_c} \frac{S+1}{S} \begin{cases} \frac{T_c}{T S} & T \ll \min(T_c, \frac{T_c}{S}), \\ \frac{1}{3}\left[1 + \left(\frac{S}{S+1}\right)\frac{T}{T_c}\right] & \frac{T_c}{S} \ll T \ll T_c, \\ \frac{T_c}{2(T_c - T)} & \frac{T_c}{T_c - T} \gg 1, \\ \frac{T_c}{T - T_c} & \frac{T_c}{T - T_c} \gg 1. \end{cases} \tag{42}$$

Note that its behavior is in agreement with the well-known relation, proposed by Koopmans et al. (2010), that the ultrafast demagnetization time scales with the ratio $\mu_{at}/T_c$ and proposed by Kazantseva et al. (2008b) and Ostler et al. (2012) that it scales as a ratio $\mu_{at}/\lambda$. As was pointed out elsewhere by Atxitia and Chubykalo-Fesenko (2011) and Atxitia et al. (2014), the complete expression involves the
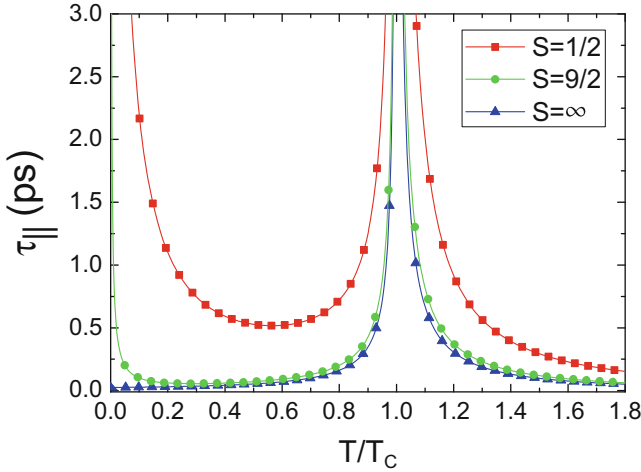
**Fig. 6** Longitudinal relaxation time (Eq. (39)) versus temperature using constant $\lambda = 0.02$, $T_c = 650$ K, and $\mu_{at} = 0.5\mu_B$ in the three spin cases with $S = 1/2$, $S = 9/2$, and $S = \infty$. The case $S = \infty$ is done by taking the limit $S \rightarrow \infty$ in Eq. (39), which is equivalent to the classical LLB equation. (Reprinted from Nieves et al. 2014)

combination of both. The two last lines in Eq. (42) describe the effect of the critical slowing down near the critical temperature. Note that the relaxation time is twice large above $T_c$ than below it. Furthermore, the relaxation time decreases with the increase of the quantum number $S$. This is due to the fact that for larger $S$ one increases the number of scattering possibilities. Furthermore, there is also an increase of the longitudinal relaxation time for low temperatures, visible especially for $S = 1/2$. This is due to the well-known fact of the freezing of fluctuations at low temperature in the quantum case and is reflected also in the fact that the perpendicular damping parameter goes to zero in the quantum case (see Eq. (27)), while it tends to a constant value in the classical case. In any case, there is an additional temperature dependence not considered above due to the fact that $\lambda(T)$ is a function of temperature itself since it contains the scattering probabilities. This dependence is specific for the scattering mechanism (see Nieves et al. 2014, Nieves 2015).

While in the LLB equation we use the longitudinal susceptibility at zero field (diverging at $T_c$), the resulting susceptibility at constant field does not diverge. In Fig. 7 we present the longitudinal relaxation time as a function of the temperature in constant applied field for the two limiting cases $S = 1/2$ and $S = \infty$. The longitudinal relaxation time was evaluated by direct integration of the qLLB equation with initial conditions $m_0 - m_e = 0.1m_e$. The longitudinal relaxation time for $S = 1/2$ is smaller in the classical case than for the quantum one, and, as expected, the maximum is displaced for larger values at larger fields.

For the two-sublattice case, it is also possible to analyze the relaxation rates of the two sublattices. However, the equations are coupled and the situation depends
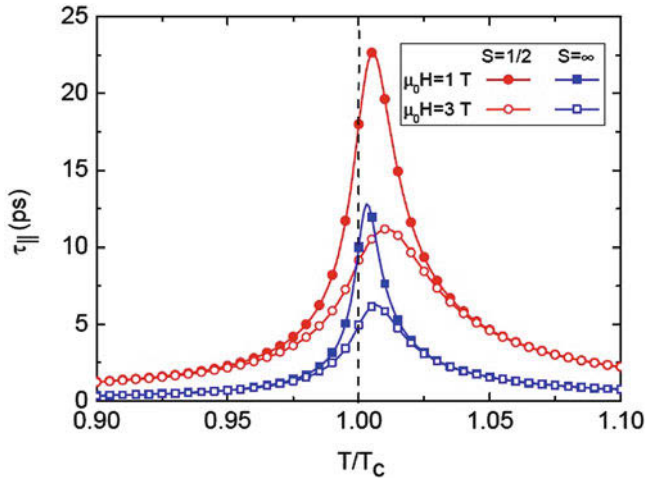
**Fig. 7** The in-field longitudinal relaxation time calculated via direct integration of the qLLB equation with small deviation from the equilibrium. The following parameters are used $T_c = 650\,$K, $\mu_{at} = 0.5\mu_B$, $\lambda = 0.02$, and zero anisotropy constant. (Reprinted from Nieves et al. 2014)
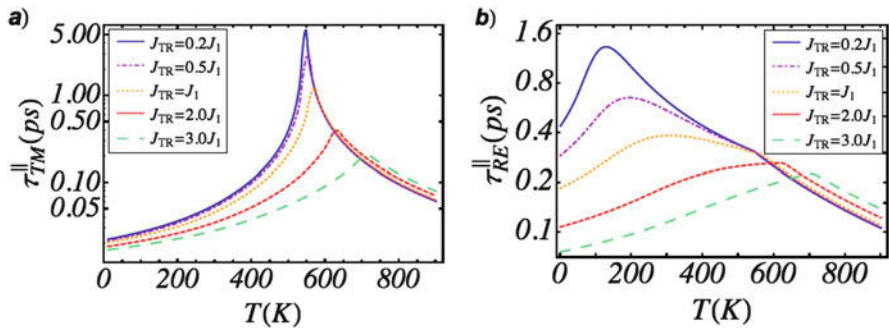


**Fig. 8** Longitudinal relaxation time of the (**a**) transition metal (TM), strong intersublattice exchange and (**b**) rare earth (RE), weak intersublattice exchange, evaluated for the $Gd_{0.25}Fe_{0.75}$ compound and different intersublattice strengths $J_{TR}$. The parameter $J_1$ corresponds to a typical (relatively weak) intersublattice strength of this material (see Suarez et al. 2015 and Ostler et al. 2011 for the values). (Reprinted from Suarez et al. 2015)

strongly on the coupling strength. Typically the divergence of the longitudinal relaxation time is suppressed by interactions (as in the case of the external field), and individual longitudinal relaxation times simply have maximum at different critical temperatures for each sublattice. An example of the longitudinal relaxation time for different intersublattice exchange strengths is presented in Fig. 8. In weakly coupled ferrimagnets, only the material with the largest exchange value slows down at the common Curie temperature (see Suarez et al. 2015).

If the coupling is not very strong and for relatively low temperatures, we can estimate the longitudinal relaxation time as defined by the exchange field acting on each sublattice as

$$\tau_v^{\|} \approx \frac{1}{2\gamma_v \lambda_v m_{v,e} H_{v,e}^{ex}}, \tag{43}$$

where

$$H_{v,e}^{ex} = \frac{\widetilde{J}_{0,v,e}}{\mu_v} m_{e,v}, \tag{44}$$

is the homogeneous exchange field evaluated at the equilibrium. Note that in this approximation, the relaxation time is independent on the sign of the coupling between sublattices (ferro or antiferro).

Interestingly that varying the initial temperature, the demagnetization speed of the initially slower material may become faster than that of the high-speed material. For example, Gd in GdFe may become faster than Fe at high initial temperatures (Atxitia et al. 2014). The temperatures at which the longitudinal relaxation times are maximum coincide for each sublattice in the case of strongly coupled alloys.

The extension of the two-sublattice LLB above the Curie temperature can be found in Nieves et al. (2015) and Nieves (2015). Here the longitudinal and the transverse damping parameters become the same and defined by the formula (Eq. (16)) for each sublattice.

## 5 Examples of Modeling Magnetization Dynamics Close to the Phase Transition

The main difference of the LLG and LLB magnetization dynamics is the presence of the longitudinal relaxation. This relaxation manifests itself close to the Curie temperature only and especially at picosecond-femtosecond timescale. In fact, close to this temperature, the magnetization can be switched by a so-called linear reversal, i.e., when the macroscopic magnetization changes its length instead of rotating the magnetization vector as it happens at low temperatures. In the intermediate region the reversal is so-called elliptical, i.e., the magnetization vector rotates together with the change of its magnitude. For the analysis of elliptical and linear reversal within the LLB model, see Kazantseva et al. (2009). The high-temperature magnetization dynamics close to $T_c$ is also characterized by the presence of the linear domain walls in which the magnetization changes its length rather than rotates (see, e.g., Hinzke et al. 2007).

To illustrate the appearance and importance of linear magnetization reversal path, we present in Fig. 9 the modeling results for the magnetization reversal time versus temperature for a magnetic nanoparticle with $S = 1/2$ (pure quantum case) and $S = \infty$ (classical case) under applied field $\mu_0 H_z = -1T$ for two different initial
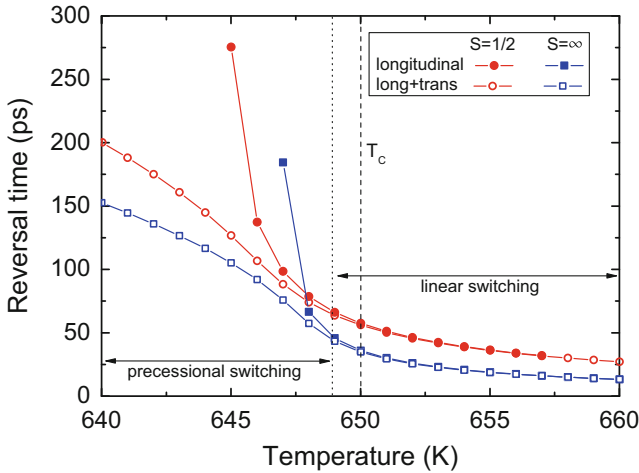
**Fig. 9** Reversal time versus temperature for a magnetic nanoparticle with magnetic moment $\mu = 0.5\mu_B$ ($\mu_B$ is the Bohr magneton), $T_c = 650$ K, and the coupling to the bath parameter $\lambda = 0.02$ under applied field $\mu_0 H_z = -1T$ modeled with the LLB equation with $S = 1/2$ and $S = \infty$. In the pure longitudinal dynamics, the initial magnetization is set to $\mathbf{m} = (0, 0, 0.2)$, and in the longitudinal plus transverse dynamics, the initial magnetization is set to $\mathbf{m} = (0.05, 0, 0.2)$. (Reprinted from Nieves et al. 2014)

conditions: (i) pure longitudinal dynamics where we put the initial magnetization of nanoparticle parallel to the field so that it cannot precess and (ii) longitudinal plus transverse dynamics where the initial magnetization is slightly tilted from the field axis. We define the reversal time as time elapsed between the initial state and the instant of time at which the magnetization begins to reverse its direction, i.e., crosses $m_z = 0$ point. One can observe that slightly below and above $T_c$, the magnetization reversal becomes completely linear, i.e., occurs by a pure change of the magnetization magnitude. However, at 5 degrees from $T_c$, the path is elliptical and at 10 degrees, it is completely precessional.

The longitudinal relaxation especially manifests itself in the laser-induced magnetization dynamics which recently has become an essential part of the novel field of opto-magnetism (Kimel et al. 2007; Kirilyuk et al. 2010). Indeed the timescale in these experiments goes down to the femtosecond range, and the energies are such that the electronic temperature often exceeds the Curie temperature of the materials. One of the main acting mechanisms has a pure heat origin (see Ostler et al. 2012), and during the femtosecond timescale, the exchange interaction energy is probed. The modeling of the ultrafast laser-induced magnetization dynamics based on the LLB equation has been successfully performed in Ni (Atxitia et al. 2010a), in FeNi (Hinzke et al. 2015), Gd (Sultan et al. 2012), and FePt (Mendil et al. 2014) with a very good agreement with experiment. For this purpose the LLB equation is coupled with the electronic temperature of the two-temperature model assuming the electron scattering mechanism.
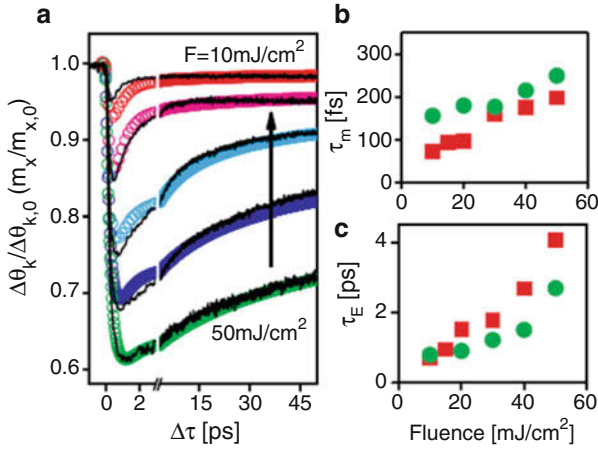
**Fig. 10** (left panel) The measured Kerr angle rotation (solid line) and the modeled magnetization dynamics (circles) for various laser fluencies $F$ on Ni. The right panels represent experimental (circles) and modeled (squares) timescales for demagnetization $\tau_m$ and recovery $\tau_E$. (Reprinted from Atxitia 2012)

The main consequence of the longitudinal relaxation is the slowing down of the first laser-induced demagnetization timescale. An example of this is illustrated in Fig. 10 where we present a comparison between the experiment performed in Atxitia et al. (2010a) on Ni and the micromagnetic modeling based on the LLB equation. The right panels represent the behavior of the demagnetization $\tau_m$ and the recovery $\tau_E$ times as a function of the laser fluency. The slowing down of the demagnetization timescale is the consequence of the slowing down of the longitudinal relaxation time, while the slowing down of the recovery timescale is the consequence of the electronic temperature dynamics. The comparison of the experimental measurements with the modeling based on the LLB has allowed to conclude the pure thermal origin of the laser-induced demagnetization. This was later confirmed by atomistic modeling and experiments on different FeCoGd thin films and dots by Ostler et al. (2012).

Another prominent example is the magnetization dynamics measured and modeled in FePt thin film with a linearly polarized laser pulse by Mendil et al. (2014). Here as the laser pulse intensity increases and the quasi-equilibrium electron temperature approaches the Curie temperature, the slowing down of the demagnetization time of FePt in the picosecond timescale is observed experimentally and in modeling. At large pulse intensity the quasi-equilibrium electron temperature stays near $T_c$, and the recovery does not take place in the 100 ps timescale. The result of the modeling is presented in Fig. 11. The experiment and LLB-based modeling with circularly polarized laser pulse can be found in John et al. (2017). The circular polarized light produces an additional effect based on the inverse Faraday one (Hertel 2005; Vahaplar et al. 2012; Battiato et al. 2014). This suppresses the slowing

**Fig. 11** Magnetization dynamics in FePt under the linearly polarized laser pulse obtained by the integration of the qLLB equation with $S = 3/2$ coupled with the two-temperature model with the parameters from Mendil et al. (2014), Nieves (2015) and for several laser intensities measured in fluence $F$. (Reprinted from Nieves 2015)
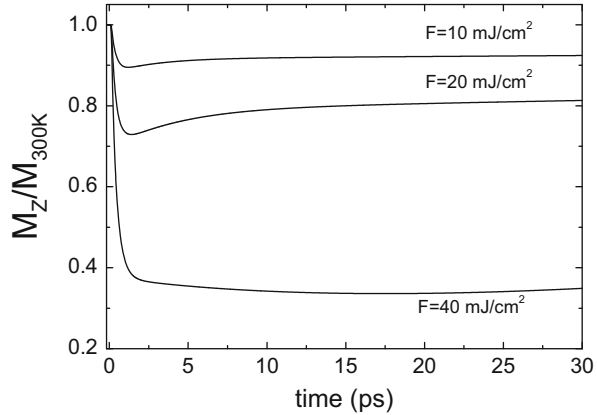
down and makes the reversal probabilities from up to down and from down to up to become non-equal. A similar effect is produced treating the inverse Faraday effects as an additional short timescale field (Nieves and Chubykalo-Fesenko 2016). In John et al. (2017) the reversal probabilities were modeled within the multiscale approach based on the DFT calculations, LLB approach, and the rate equations. The modeling results explained the recently observed by Lambert et al. (2014) switching of FePt granular media by circularly polarized pulses.

The two-sublattice LLB approach was successfully used to explain the temperature dependence of the frequencies and damping parameter of ferro- and antiferromagnetic modes in GdFe ferrimagnet (Schlickeiser et al. 2012) and to explain the different demagnetization speeds in Fe and Ni in permalloy (Hinzke et al. 2015).

Furthermore the LLB approach is currently extensively used as a framework for modeling of heat-assisted magnetic recording (McDaniel 2012; Vogler et al. 2014, 2016). It is especially useful for a coarse-grained modeling when one magnetic grain can be represented as the LLB macrospin (Vogler et al. 2014, 2016).

Recently several extensions of the LLB model also appeared, to mention the introduction of the spin-torque effects into the LLB model by Schieback et al. (2009) or Janda et al. (2017), of the colored noise (Atxitia 2012) or the self-consistent modeling of magnetization and temperature dynamics (Nieves et al. 2016) useful for modeling of magnetocaloric effect and magnetic hyperthermia.

## 6   Conclusions

The LLB approach is a valuable micromagnetic framework to model magnetization dynamics close to $T_c$ and in general where rapid temperature changes appear. First, it correctly reproduces the temperature dependence of relaxational parameters and is in agreement with atomistic modeling. Second, it has proved its viability in modeling of several important research areas such as the heat-assisted magnetic

recording, spin caloritronics, and ultrafast laser-induced magnetization dynamics and was validated via the experimental measurements. It is an important part of the multiscale framework, especially when the model should include the temperature effects. The introduction of the quantum LLB equation allows one to pass directly from ab initio modeling to the micromagnetic one without going through the atomistic part and avoiding the classical temperature dependencies of the micromagnetic parameters as opposed to the more correct quantum dependencies. This, however, needs future investigation on the relevant scattering mechanisms and their spin-flip probabilities. The generalization of the LLB equation for the case of alloys allows one to describe separately the dynamics of each species, in agreement with experimental predictions that these dynamics are different on the ultrashort timescale.

# References

Asselin P, Evans RFL, Barker J, Chantrell RW, Yanes R, Chubykalo-Fesenko O, Hinzke D, Nowak U (2010) Constrained Monte Carlo method and calculation of the temperature dependence of magnetic anisotropy. Phys Rev B 82:054415

Atxitia U (2012) Modeling of ultrafast laser-induced magnetization dynamics within the Landau-Lifshitz-Bloch approach. Ph.D. thesis, Instituto de Ciencia de Materiales de Madrid (ICMM) – Universidad Autónoma de Madrid

Atxitia U, Chubykalo-Fesenko O (2011) Ultrafast magnetization dynamics rates within the Landau-Lifshitz-Bloch model. Phys Rev B 84:144414

Atxitia U, Chubykalo-Fesenko O, Kazantseva N, Hinzke D, Nowak U, Chantrell RW (2007) LLB-micromagnetic modelling of laser-induced magnetisation dynamics. Appl Phys Lett 91:232507

Atxitia U, Chubykalo-Fesenko O, Walowski J, Mann A, Münzenberg M (2010a) Evidence for thermal mechanisms in laser-induced femtosecond spin dynamics. Phys Rev B 81:174401

Atxitia U, Hinzke D, Chubykalo-Fesenko O, Nowak U, Kachkachi H, Mryasov ON, Evans RF, Chantrell RW (2010b) Multiscale modeling of magnetic materials: temperature dependence of the exchange stiffness. Phys Rev B 82:134440

Atxitia U, Nieves P, Chubykalo-Fesenko O (2012) Landau-Lifshitz-Bloch equation for ferrimagnetic materials. Phys Rev B 86:104414

Atxitia U, Barker J, Chantrell RW, Chubykalo-Fesenko O (2014) Controlling the polarity of the transient ferromagneticlike state in ferrimagnets. Phys Rev B 89:224421

Atxitia U, Hinzke D, Nowak U (2016) Fundamentals and applications of the Landau-Lifshitz-Bloch equation. J Phys D Appl Phys 50:033003

Battiato M, Barbalinardo G, Oppeneer PM (2014) Phys Rev B 89:014413

Beaurepaire E, Merle JC, Daunois A, Bigot JY (1996) Ultrafast spins dynamics in ferromagnetic nickel. Phys Rev Lett 76:4250

Bloch F (1946) Nuclear induction. Phys Rev 70:460

Blum K (1981) Density matrix theory and applications. Plenum Press, New York

Brown WF (1963a) Micromagnetics. Wiley, New York

Brown WF (1963b) Thermal fluctuations of a single-domain particle. Phys Rev 130:1677

Chantrell RW, Wongsam M, Schrefl T, Fidler J (2001) Micromagnetics I: basic principles. In: Buschow KHJ, Cahn RW, Flemings MC, Ilschner B, Kramer EJ, Mahajan S (eds) Encyclopedia of materials: science and technology. Elsevier, Amsterdam

Chubykalo O, Hannay JD, Wongsam MA, Chantrell RW, González JM (2002) Langevin dynamic simulation of spin waves in a micromagnetic model. Phys Rev B 65:184428

Chubykalo O, Smirnov-Rueda R, Wongsam MA, Chantrell RW, Nowak U, González JM (2003) Brownian dynamics approach to interacting magnetic moments. J Magn Magn Mat 266:28

Chubykalo-Fesenko O, Nowak U, Chantrell RW, Garanin D (2006) Dynamic approach for micromagnetics close to the Curie temperature. Phys Rev B 74:094436

Coey J (2009) Magnetism and magnetic materials. Cambridge University Press, New York

Eriksson O, Bergman A, Bergqvist L, Hellsvik J (2017) Atomistic spin dynamics foundations and applications. Oxford University Press, New York

Evans RFL, Hinzke D, Atxitia U, Nowak U, Chantrell RW, Chubykalo-Fesenko O (2012) Stochastic form of the Landau-Lifshitz-Bloch equation. Phys Rev B 85:014433

Evans RFL, Fan WJ, Chureemart P, Ostler TA, Ellis MOA, Chantrell RW (2014) Atomistic spin model simulations of magnetic nanomaterials. J Phys Cond Mat 26:103202

Fidler J, Schrefl T (2000) Micromagnetic modelling—the current state of the art. J Phys D Appl Phys 33:R135

Garanin DA (1991) Generalized equation of motion for a ferromagnet. Phys A 172:470

Garanin DA (1997) Fokker-Planck and Landau-Lifshitz-Bloch equation for classical ferromagnets. Phys Rev B 55:3050

Garanin DA (2012) Density matrix equation for a bathed small system and its application to molecular magnets. Adv Chem Phys 147:213

Garanin DA, Chubykalo-Fesenko O (2004) Thermal fluctuations and longitudinal relaxation of single-domain magnetic particles at elevated temperatures. Phys Rev B 70:212409

Garanin DA, Ishtchenko VV, Panina LV (1990) Dynamics of an ensemble of single-domain magnetic particles. Theor Math Phys 82:169

García-Palacios JL, Lázaro FJ (1998) Langevin-dynamics study of the dynamical properties of small magnetic particles. Phys Rev B 58:14937

Gilbert T (2004) A phenomenological theory of damping in ferromagnetic materials. IEEE Trans Magn 40:6

Greaves SJ, Muraoka H, Kanai Y (2015) Modelling of heat assisted magnetic recording with the landau-lifshitz-bloch equation and brillouin functions. J Appl Phys 117:17C505

Grinstein G, Koch RH (2003) Coarse graining in micromagnetics. Phys Rev Lett 90:207201

Hertel R (2005) Theory of the inverse faraday effect in metals. J Magn Magn Mater 202:L1–L4

Hinzke D, Nowak U (2011) Domain wall motion by the magnonic spin Seebeck effect. Phys Rev Lett 107:027205

Hinzke D, Nowak U, Garanin DA (2000) Uniform susceptibility of classical antiferromagnets in one and two dimensions in a magnetic field. Euro Phys J B 16:435

Hinzke D, Nowak U, Mryasov ON, Chantrell RW (2007) Orientation and temperature dependence of domain wall properties in FePt. Appl Phys Lett 90:082507

Hinzke D, Atxitia U, Carva K, Nieves P, Fesenko-Chubykalo O, Oppeneer P, Nowak U (2015) Multiscale modeling of ultrafast element specific magnetization dynamics of ferromagnetic alloys. Phys Rev B 92:054412

Janda T, Roy P, Otxoa R, Soban Z, Ramsay A, Irvine A, Trojanek F, Surynek M, Campion R, Gallagher B, Jungwirth T, Nemec P, Wunderlich J (2017) Inertial displacement of a domain wall excited by ultra-short circularly polarized laser pulses. Nat Commun 8:15226

John R, Berrita M, Hinzke D, Muller C, Santos T, Ulrichs H, Nieves P, Mondal R, Walowski J, Chubykalo-Fesenko O, McCord J, Oppenneer P, Nowak U, Muzenberg M (2017) Magnetization switching of FePt nanoparticle recording medium by femtosecond laser pulses. Sci Rep 7:4114

Kachkachi H, Garanin D (2001) Magnetic free energy at elevated temperatures and hysteresis of magnetic particles. Phys A 291:485–500

Kazantseva N (2008) Dynamic response of the magnetisation to picosecond heat pulses. Ph.D. thesis, University of York

Kazantseva N, Hinzke D, Nowak U, Chantrell RW, Chubykalo-Fesenko O (2007) Atomistic models of ultrafast reversal. Phys Stat Sol 244:4389

Kazantseva N, Hinzke D, Nowak U, Chantrell RW, Atxitia U, Chubykalo-Fesenko O (2008a) Towards multiscale modelling of magnetic materials: simulations of FePt. Phys Rev B 77:184428

Kazantseva N, Nowak U, Chantrell RW, Hohlfeld J, Rebei A (2008b) Slow recovery of the magnetisation after a sub-picosecond heat-pulse. Europhys Lett 81:27004

Kazantseva N, Hinzke D, Chantrell R, Nowak U (2009) Linear and elliptical magnetization reversal close to the curie temperature. Europhys Lett 86:27006

Kimel AV, Kirilyuk A, Rasing T (2007) Femtosecond opto-magnetism: ultrafast laser manipulation. Laser Photonics Rev 1:275–287

Kirilyuk A, Kimel AV, Rasing T (2010) Ultrafast optical manipulation of magnetic order. Rev Mod Phys 82:2731

Koopmans B, Ruigrok JJM, Longa FD, de Jonge WJM (2005) Unifying ultrafast magnetization dynamics. Phys Rev Lett 95:267207

Koopmans B, Malinowski G, Longa FD, Steiauf D, Fähnle M, Roth T, Cinchetti M, Aeschlimann M (2010) Explaining the paradoxical diversity of ultrafast laser-induced demagnetization. Nat Mat 9:259–265

Lambert CH, Mangin S, Varaprasad BSDCS, Takahashi YK, Hehn M, Cinchetti M, Malinowski G, Hono K, Fainman Y, Aeschlimann M, Fullerton EE (2014) All-optical control of ferromagnetic thin films and nanostructures. Science 345:1337

Landau DL, Lifshitz E (1935) On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. Phys Z Sowjetunion 8:153

Lyberatos A, Chantrell RW (1993) Thermal fluctuations in a pair of magnetostatistically coupled particles. J Appl Phys 73:6501

McDaniel TW (2012) Application of Landau-Lifshitz-Bloch dynamics to grain switching in heat-assisted magnetic recording. J Appl Phys 112:013914

Mendil J, Nieves P, Chubykalo-Fesenko O, Walowski J, Santos T, Pisana S, Münzenberg M (2014) Resolving the role of femtosecond heated electrons in ultrafast spin dynamics. Sci Rep 4:3980

Moreno R, Evans R, Khmelevskyi S, Munoz M, Chantrell R, Chubykalo-Fesenko O (2016) Temperature-dependent exchange stiffness and domain wall width in Co. Phys Rev B 94:104433

Nakatani Y, Uesaka Y, Hayashi N, Fukushima H (1997) Computer simulation of thermal fluctuation of fine particle magnetization based on Langevin equation. J Magn Magn Mat 168:347

Nieves P (2015) Micromagnetic models for high-temperature magnetization dynamics. Ph.D. thesis, Instituto de Ciencia de Materiales de Madrid (ICMM) – Universidad Autónoma de Madrid

Nieves P, Chubykalo-Fesenko O (2016) Modeling of ultrafast heat- and field-assisted magnetization dynamics in FePt. Phys Rev Appl 5:014006

Nieves P, Serantes D, Atxitia U, Chubykalo-Fesenko O (2014) The quantum Landau-Lifshitz-Bloch equation and its comparison with the classical case. Phys Rev B 90:104428

Nieves P, Atxitia U, Chantrell RW, Chubykalo-Fesenko O (2015) The classical two sublattice Landau-Lifshitz-Bloch equation at all temperatures. Low Temp Phys 41:949

Nieves P, Serantes D, Chubykalo-Fesenko O (2016) Self-consistent description of spin-phonon dynamics in ferromagnets. Phys Rev B 94:014409

Nieves P, Arapan S, Schrefl T, Cuesta-Lopez S (2017) Atomistic spin dynamics simulations of the MnAl $\tau$-phase and its antiphase boundary. Phys Rev B 96:224411

Ostler TA, Evans RFL, Chantrell RW, Atxitia U, Chubykalo-Fesenko O, Radu I, Abrudan R, Radu F, Tsukamoto A, Itoh A, Kirilyuk A, Rasing T, Kimel A (2011) Crystallographically amorphous ferrimagnetic alloys: comparing a localized atomistic spin model with experiments. Phys Rev B 84:024407

Ostler T, Barker J, Evans R, Chantrell R, Atxitia U, Chubykalo-Fesenko O, Moussaoui SE, Guyader LL, Mengotti E, Heyderman L, Nolting F, Tsukamoto A, Itoh A, Afanasiev D, Ivanov

B, Kalashnikova A, Vahaplar K, Mentink J, Kirilyuk A, Rasing T, Kimel A (2012) Ultrafast heating as a sufficient stimulus for magnetization reversal in a ferrimagnet. Nat Commun 3:666

Schieback C, Hinzke D, Kläui M, Nowak U, Nielaba P (2009) Temperature dependence of the current-induced domain wall motion from a modified Landau-Lifshitz-Bloch equation. Phys Rev B 80:214403

Schlickeiser F, Atxitia U, Wienholdt S, Hinzke D, Chubykalo-Fesenko O, Nowak U (2012) Temperature dependence of the frequencies and effective damping parameters of ferrimagnetic resonance. Phys Rev B 86:214416

Schlickeiser F, Ritzmann U, Hinzke D, Nowak U (2014) Role of entropy in domain wall motion in thermal gradients. Phys Rev Lett 113:097201

Scholz W, Schrefl T, Fidler J (2001) Micromagnetic simulation of thermally activated switching in fine particles. J Magn Magn Mat 233:296

Skubic B, Hellsvik J, Nordström L, Eriksson O (2008) A method for atomistic spin dynamics simulations: implementation and examples. J Phys Condens Matter 20:315203

Suarez O, Nieves P, Laroze D, Altbir D, Chubykalo-Fesenko O (2015) The ultra-fast relaxation rates and reversal time in disordered ferrimagnets. Phys Rev B 92:144425

Sultan M, Atxitia U, Melnikov A, Chubykalo-Fesenko O, Bovensiepen U (2012) Electron- and phonon-mediated ultrafast magnetization dynamics of Gd(0001). Phys Rev B 85:184407

Takano K, Jin E, Zhou D, Maletzky T, Smyth J, Dovek M (2011) Thermo-dynamic magnetisation model of thermally assisted magnetic recording by Landau-Lifshitz-Bloch equation. J Magn Soc Jpn 35:431

Uchida K, Takahashi S, Harii K, Ieda J, Koshibae W, Ando K, Maekawa S, Saitoh E (2008) Observation of the spin Seebeck effect. Nature 455:778

Vahaplar K, Kalashnikova AM, Kimel A, Hinzke D, Nowak U, Chantrell R, Tsukamoto A, Itoh A, Kirilyuk A, Rasing T (2009) Ultrafast path for optical magnetization reversal via a strongly nonequilibrium state. Phys Rev Lett 103:117201

Vahaplar K, Kalashnikova AM, Kimel A, Gerlach S, Hinzke D, Nowak U, Chantrell R, Tsukamoto A, Itoh A, Kirilyuk A, Rasing T (2012) All-optical magnetization reversal by circularly polarized laser pulses: Experiment and multiscale modeling. Phys Rev B 85:104402

Vogler C, Abert C, Bruckner F, Suess D (2014) Landeu-Lifshits-Bloch equation for exchange-coupled grains. Phys Rev B 90:214431

Vogler C, Abert C, Bruckner F, Suess D, Praetorius D (2016) Areal density optimizations for heat-assisted magnetic recording of high-density media. J Appl Phys 120:223903

Vogler C, Abert C, Bruckner F, Suess D (2018) Stochastic ferrimagnetic Landau-Lifshitz-Bloch equation for finite magnetic structures. arXiv:180401724

Xu L, Zhang S (2012) Magnetization dynamics at elevated temperatures. Phys E 45:72

# Density Functional Theory for Magnetism and Magnetic Anisotropy

# 42

Gustav Bihlmayer

# Contents

## Abstract

Density functional theory and its application for the simulation of magnetic properties of condensed matter is introduced. This includes vector-spin density functional theory for the evaluation of spin-spin interactions and relativistic extensions to capture effects like the magnetocrystalline anisotropy. The role of the different approximations to the exchange-correlation functional, e.g., the local density approximation, or the generalized gradient approximation, is investigated, showing successes and limitations of the present functionals. Special techniques to determine, e.g., the magnetic ground state or finite temperature properties based on density functional theory are shortly discussed.

G. Bihlmayer (✉)
Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany
e-mail: g.bihlmayer@fz-juelich.de

# 1 Introduction

Being a material-specific theory, density functional theory (DFT) has gained enormous popularity in the last few decades for the investigation of structural, electronic, or optical properties of condensed matter and is valued as "approximate practical method" in many other areas as well (Jones 2015). In the field of magnetism the value of DFT calculations is widely accepted: only few input parameters have to be known (the chemical composition and, maybe, some information on the crystal structure) to calculate ground state properties like the magnetic order (ferro- or antiferromagnetic, non-collinear structures etc.), spin- and orbital magnetic moments, magnetic interaction parameters (e.g., the spin-stiffness), or the easy axis (or plane) of magnetization. These "ab-initio" parameters can further be used to determine properties beyond the ground state, for example spin-wave spectra or ordering temperatures. In many cases, the results are of excellent quality, in particular in the field of itinerant magnets DFT results usually have predictive power.

To illustrate these statements, magnetic moments and ordering temperatures of several Co-based Heusler compounds are compiled in Table 1 (Thoene et al. 2009). The structure of these compounds consists of four interpenetrating face-centered cubic (fcc) lattices where the atoms of the formula $Co_2YZ$ are located. Here, $Y$ is another transition metal element, while $Z$ is from the $p$-block of the main group. The DFT calculations give access not only to the total magnetic moment in the unit cell, $M_{tot}$, which is in good agreement with the experimental values, but also to locally resolved quantities. E.g., one can see a considerable variation of the Co moments for the listed $Co_2YZ$ structures, depending on their neighboring atoms. With increasing total moment also the Curie-temperatures rises in the shown series, a trend that is also confirmed experimentally.

Despite the many successes of DFT in the field of magnetism, one has to be aware of possible complications that can arise due to the specific nature of some materials. This can be e.g., correlation effects that are difficult to capture in DFT calculations (Galler et al. 2015) or zero-point fluctuations that have to be accounted for (Ibáñez Azpiroz et al. 2016). However, also in these cases DFT can be used as a starting point to explore the magnetic properties of materials.

All specific examples shown in this chapter are bulk magnets that can be reliably calculated with different basis sets, e.g., planewaves, augmented plane

**Table 1** Magnetic moment, $M$, in $\mu_B$ per site (Co or $Y$) or unit cell (tot) and Curie temperature ($T_C$, in Kelvin) of several Heusler compounds: The moments and exchange interactions were obtained in DFT, while $T_C$ is calculated from these DFT results in the mean-field approximation. Experimental values are listed for comparison. (From Thoene and coworkers 2009)

| $Co_2YZ$ | $M_{Co}^{calc}$ | $M_Y^{calc}$ | $M_{tot}^{calc}$ | $M_{tot}^{exp}$ | $T_C^{calc}$ | $T_C^{exp}$ |
|---|---|---|---|---|---|---|
| $Co_2TiAl$ | 0.48 | −0.08 | 0.85 | 0.74 | 148 | 134 |
| $Co_2VGa$ | 0.89 | 0.20 | 1.95 | 1.92 | 343 | 352 |
| $Co_2MnGa$ | 0.79 | 2.80 | 4.29 | 4.11 | 698 | 685 |
| $Co_2FeSi$ | 1.30 | 2.76 | 5.34 | 6.00 | 1134 | 1100 |

waves, muffin-tin orbitals, etc. assuming three-dimensional periodicity. For other material classes (like molecular magnets or adatoms on surfaces) other basis sets (localized orbitals) or formalisms (Green function methods) might be better suited. Relativistic calculations (to obtain e.g., spin-orbit coupling effects) or problems dealing with electron-nucleus interaction might profit from an all-electron treatment of the DFT problem, while for obtaining other quantities (spin moments, scalar exchange interactions) a pseudopotential treatment can be sufficient. In most cases, however, the diversity of available DFT implementations will make it easy to find a suitable solution for a specific problem.

## 2    Methodological Framework

To calculate the electronic structure of an atom, a molecule or a solid, it is necessary to solve the Schrödinger- or Dirac-equation that governs the behavior of the electrons in the system. This behavior is encoded in the wavefunction, $\Psi$, that is a complex quantity in $3N$-dimensional space if $N$ is the number of electrons. Even for a moderate amount of particles (e.g., 26 electrons in an Fe atom) this many-particle wavefunction is impossible to handle numerically. If spin comes into play, or a time-dependence has to be included, the problem even worsens. Therefore, a straightforward solution is – in most cases – out of reach.

The key idea of density functional theory is to work, instead of the many-body wavefunction, with the density, $n(\mathbf{r})$, as the basic variable. The latter can be obtained by the former by integration over all but one spatial variables:

$$n(\mathbf{r}) = N \int d\mathbf{r}_2 \dots \int d\mathbf{r}_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N). \tag{1}$$

Surprisingly, the total energy, $E$, turns out to be a unique functional of this density and this functional is stationary with respect to variations of $n(\mathbf{r})$. This theorem was worked out by Hohenberg and Kohn (1964), together with the observation that in a non-degenerate ground state the many-body wavefunction $\Psi$ that describes electrons in an external potential, $v(\mathbf{r})$ (caused by e.g., nuclei), is uniquely determined by the particle density distribution $n(\mathbf{r})$ (The charge density is obtained by multiplying $n$ with the electron's charge $-e$. In atomic units this is normalized to unity and in the following $e$ is not written explicitly, unless in the context of some relativistic terms.). The energy functional can be written as

$$E[n(\mathbf{r})] = \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})], \tag{2}$$

where the first term on the right side describes the Coulomb interaction of the density with the external potential and the functional $F$ captures the kinetic energy of the electrons and their mutual Coulomb repulsion. Since Hohenberg and Kohn could show that the density determines $v(\mathbf{r})$ to within a constant, all terms in Eq. (2) can be determined for a given $n(\mathbf{r})$, provided that $F$ is known.

It was the achievement of Kohn and Sham (1965) to find a representation of this functional that allowed to split it into large, known parts and a small, unknown remainder. Their idea was to consider a fictitious system of independent electrons that have the same density, $n(\mathbf{r})$, as the true electron system: Then, the kinetic energy, $T_0$, and the Coulomb interaction of these electrons can be separated out:

$$F[n(\mathbf{r})] = T_0[n(\mathbf{r})] + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{xc}}[n(\mathbf{r})] \qquad (3)$$

and the remaining so-called exchange-correlation (XC) functional, $E_{\text{xc}}[n(\mathbf{r})]$, is just a small correction to be determined. In reality, the movement of the electrons is correlated due to their Coulomb interaction, not independent, and $E_{\text{xc}}$ has to account for this difference. The exchange energy, arising from the fact that the density is derived from a Slater determinant of wavefunctions that will be defined below, is also contained in $E_{\text{xc}}$.

With this decomposition, it is possible to evaluate the (large) term $T_0$, assuming that the fictitious independent electrons are described by wavefunctions $\phi$ so that $n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2$. Hereby, an Aufbau principle is used to occupy the $\phi_i$, ensuring that the total wavefunction of the Kohn-Sham system satisfies the Pauli principle. The kinetic energy of the electrons with the mass $m$ is given by $T_0 = -\frac{\hbar^2}{2m} \sum_{i=1}^{N} \phi_i^*(\mathbf{r})\nabla^2 \phi_i(\mathbf{r})d\mathbf{r}$. Now, requiring that $E[n(\mathbf{r})]$ is stationary with respect to variations of the ground state density and requiring particle conservation, the resulting Euler-Lagrange equation can be recast in the form of an effective single particle Schrödinger equation, the Kohn-Sham equation:

$$\left[ -\frac{\hbar^2}{2m}\nabla^2 + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' + v(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) . \qquad (4)$$

Here, the exchange correlation potential, $V_{\text{xc}}$, is the derived as $\frac{\delta E_{\text{xc}}}{\delta n(\mathbf{r})}$, the $\varepsilon_i$ are strictly speaking just Lagrange parameters that follow from the normalization condition; In practice they are often used as "single particle energies" to describe spectral properties. Equation (4) has to be solved self-consistently since it contains the density (both explicitly and implicitly via the exchange correlation potential) that depends on the single particle wavefunctions. Nevertheless, the numerical effort is rather moderate, comparable to the well-known Hartree equation, i.e., it scales normally with the third power of the system size. In particular, this is much faster than wave-function based methods (Friesner 2005).

## 2.1 Vector-Spin DFT

Since the magnetic ground-state properties are uniquely determined by the many-body wavefunction that is in turn determined by the density, in principle DFT allows finding out the magnetic order, magnetic moments etc., provided that a method is known to extract these properties from the density. In practice, however, it turns out
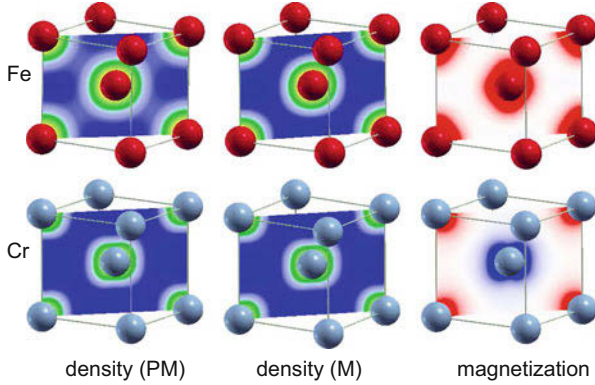
**Fig. 1** (110) cuts through the density and magnetization of bcc Fe and bcc Cr. The left and middle columns show the densities in the nonmagnetic (NM) and magnetic (M) state, respectively, where blue color indicated low density and green/yellow high density. In the right column the magnetization is plotted, where red (blue) color encodes positive (negative) values. The plots were obtained using the FLEUR-code for the DFT calculations (Kurz et al. 2004) and XCrysDen for plotting (Kokalj 2003)

to be easier to extend DFT in a way that explicitly includes the spin-density, $s(\mathbf{r})$, in the formalism. This has the additional advantage that it gives a handle to calculate not only the magnetic ground state, but also other, metastable, magnetic orders. E.g., it is possible to initialize a DFT calculation of body-centered cubic (bcc) Fe such that ferromagnetic (FM), antiferromagnetic (AFM) or nonmagnetic (NM) solutions and their relative stability are obtained. This turns out to be very useful to access the strength of magnetic interactions in a system.

As an example, in Fig. 1 cuts through the (spin-) density of Fe and Cr are shown in the NM state and in the magnetic ground state that is FM for Fe and AFM for Cr. Clearly, the density alone is not very sensitive to variations of the magnetic state, although for FM Fe a small decrease of hybridization, which is due to the spin-splitting of the states, can be seen in the images. However, in both cases only the spin-density gives a clear picture of the magnetic order.

The spin-dependent version of DFT works with spinor wavefunctions to define the spin density $\mathbf{s}(\mathbf{r})$:

$$\mathbf{s}(\mathbf{r}) = \left\langle \phi(\mathbf{r}) | \underline{\boldsymbol{\sigma}} | \phi(\mathbf{r}) \right\rangle ; \quad \phi(\mathbf{r}) = \begin{pmatrix} \phi_\uparrow(\mathbf{r}) \\ \phi_\downarrow(\mathbf{r}) \end{pmatrix} . \tag{5}$$

(The magnetization density is obtained by multiplying $-\mathbf{s}$ with the Bohr magneton, $\mu_B = \frac{e\hbar}{2mc}$.) Here, the Pauli matrices (underlined symbols denote $2 \times 2$ matrices) are given as:

$$\underline{\sigma_x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \underline{\sigma_y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \underline{\sigma_z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{6}$$

With this, a density matrix can be introduced that is composed of a scalar and a vectorial part, corresponding to the particle and spin density:

$$\underline{n}(\mathbf{r}) = \frac{1}{2} \left( n(\mathbf{r})\underline{I} + \underline{\boldsymbol{\sigma}} \cdot \mathbf{s}(\mathbf{r}) \right) = \frac{1}{2} \begin{pmatrix} n(\mathbf{r}) + s_z(\mathbf{r}) & s_x(\mathbf{r}) - i s_y(\mathbf{r}) \\ s_x(\mathbf{r}) + i s_y(\mathbf{r}) & n(\mathbf{r}) - s_z(\mathbf{r}) \end{pmatrix}, \quad (7)$$

where $\underline{I}$ denotes a $2 \times 2$ unit matrix. It is easy to see that the components of the density matrix are given by $n_{\alpha\beta}(\mathbf{r}) = \phi_\alpha^*(\mathbf{r})\phi_\beta(\mathbf{r})$ where $\alpha, \beta$ are spin labels. Analogously, a potential matrix, denoted as $\underline{v}(\mathbf{r})$, can be written in terms of a scalar potential, $v$, and magnetic field, $\mathbf{B}(\mathbf{r})$:

$$\underline{v}(\mathbf{r}) = v(\mathbf{r})\underline{I} - \mu_B \underline{\boldsymbol{\sigma}} \cdot \mathbf{B}(\mathbf{r}) . \quad (8)$$

In terms of these quantities, von Barth and Hedin (1972) extended DFT to spin-polarized systems. They derived an analog to Eq. (4) in the form:

$$\left[ \left( -\frac{\hbar^2}{2m}\nabla^2 + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right) \underline{I} + \underline{v}(\mathbf{r}) + \underline{V}_{\mathrm{xc}}(\mathbf{r}) \right] \begin{pmatrix} \phi_i^\uparrow(\mathbf{r}) \\ \phi_i^\downarrow(\mathbf{r}) \end{pmatrix} = \varepsilon_i \begin{pmatrix} \phi_i^\uparrow(\mathbf{r}) \\ \phi_i^\downarrow(\mathbf{r}) \end{pmatrix}, \quad (9)$$

where $\underline{V}_{\mathrm{xc}}$ is now defined as the functional derivative of the exchange-correlation energy with respect to the density matrix. If $\underline{v}(\mathbf{r})$ and $\underline{V}_{\mathrm{xc}}(\mathbf{r})$ are diagonal matrices, Eq. (9) clearly decomposes into two equations of type (4) for spin-up and spin-down wavefunctions.

Naturally the quality of the results obtained with Eq. (9) depends heavily on the approximation for $\underline{V}_{\mathrm{xc}}$. In the local density approximation (LDA) or, for the spin-polarized case, the local spin density approximation (LSDA), the exchange correlation energy is assumed to be of the form

$$E_{\mathrm{xc}}^{LSDA} = \int n(\mathbf{r})\varepsilon_{\mathrm{xc}} \left[ n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r}) \right] d\mathbf{r}, \quad (10)$$

where $\varepsilon_{\mathrm{xc}}$ is the exchange correlation energy density of the homogeneous, spin-polarized electron gas. Already before the formulation of DFT several approximations have been known for the exchange energy of the homogeneous electron gas, e.g., from the Thomas-Fermi-Dirac theory (Dirac 1930) or derived from Hartree-Fock theory (Slater 1951). In all these cases it was observed that $\varepsilon_{\mathrm{x}} \propto [n(\mathbf{r})]^{1/3}$ and, surprisingly, the first LDA functionals based on this approximations performed rather well.

Although modern exchange correlation functionals are more sophisticated, many properties of spin-polarized DFT calculations can be studied already in the simple LSDA form. Suppose, there is a collinear magnet with the orientation of the magnetization in $z$-direction (actually, the Hamiltonian in Eq. (9) is invariant under spin-rotations). Then, the density matrix is diagonal and $\underline{V}_{\mathrm{xc}}$ has only two terms: $V_{\mathrm{xc}}^{\uparrow\uparrow} = \frac{\delta E_{\mathrm{xc}}}{\delta n_{\uparrow\uparrow}} \propto \left[ n_{\uparrow\uparrow}(\mathbf{r}) \right]^{1/3}$ and $V_{\mathrm{xc}}^{\downarrow\downarrow} = \frac{\delta E_{\mathrm{xc}}}{\delta n_{\downarrow\downarrow}} \propto \left[ n_{\downarrow\downarrow}(\mathbf{r}) \right]^{1/3}$. This means that Eq. (9)

**Table 2** Magnetic moments (in $\mu_B$ per atom) of ferromagnetic elements in the bulk. The experimentally determined total magnetization, $M_{\text{tot.}}$, consists of spin- and orbital moment contributions. The LSDA results for Fe, Co and Ni are taken from Moruzzi et al. (1978), the GGA values of the magnetic moments from Shallcross and coworkers (2005), the experimental values are quoted from Trygg et al. (1995). The calculated Gd moments are from Kurz et al. (2002), the experimental one is taken from White and coworkers (1975)

| Property | Source | Fe (bcc) | Co | Ni (fcc) | Gd (hcp) |
|---|---|---|---|---|---|
| $M_{\text{spin}}$ | LSDA | 2.15 | 1.56 (fcc) | 0.59 | 7.63 |
| $M_{\text{spin}}$ | GGA | 2.22 | 1.62 (fcc) | 0.62 | 7.65 |
| $M_{\text{spin}}$ | Experiment | 2.13 | 1.52 (hcp) | 0.57 | |
| $M_{\text{tot.}}$ | Experiment | 2.21 | 1.66 (hcp) | 0.62 | 7.63 |

consists of two equations, one for $\phi_\uparrow$ and one for $\phi_\downarrow$ that are identical if $n_{\uparrow\uparrow} = n_{\downarrow\downarrow}$, i.e., if the charge density is not spin-polarized. If a self-consistent calculation starts from a non-magnetic density it remains in this state, even if a spin-polarized solution gives a lower total energy. This shows that the spin-polarized version of DFT can lead to metastable solutions and it is necessary to break the symmetry of the spin-channels to arrive at a magnetic ground state.

Some DFT results for ferromagnetic elements are shown in Table 2, where both the LSDA moments are listed and values obtained within the generalized gradient approximation (GGA) to the exchange-correlation potential. Although the more sophisticated GGA functionals usually lead to better predictions of the atomic structure or lattice constants, the magnetic properties are in most cases not improved (Singh and Ashkenazi 1992). Systematic studies of magnetic properties with modern XC potentials are rather scarce, but it seems that at least in some cases care has to be taken (Koller et al. 2011). This is not surprising, since most efforts are devoted to improving the description of structural parameters or binding energies in these functionals and often the performance with respect to magnetic properties is not tested. There are developments of new functionals that can, in principle, improve also the description of non-collinear magnetism (Eich and Gross 2013), however with considerably increased numerical efforts.

## 2.2   Spin-Orbit Coupling

It was already mentioned in the last section that the Hamiltonian in Eq. (9) has several shortcomings: Since it is derived from the Schrödinger equation, there is no spin-dependent term in this Hamiltonian; Spin enters just through the wavefunction and the Pauli-principle that is encoded in the exchange part of the XC potential. This is enough to find out that Fe is ferromagnetic and has a (spin-) moment of $2.2\,\mu_B$, but this moment has no preferential direction. The concept of easy and hard magnetic directions enters only if further, spin-dependent terms are added to the Hamiltonian. Thinking in terms of a general interaction between two classical spins, $\mathbf{S}$, of the form $\mathbf{S}_\nu \underline{J}_{\nu\nu'} \mathbf{S}_{\nu'}$, where $\underline{J}$ is a $3 \times 3$ matrix, only the trace of $\underline{J}$ can be accessed by

Eq. (9). This gives rise to Heisenberg-type interactions of the form $J_{\nu\nu'}\mathbf{S}_\nu \cdot \mathbf{S}_{\nu'}$ that are in many cases responsible for the magnetic order and, for $\nu = \nu'$, it describes the intra-atomic exchange interaction that is responsible for the formation of the magnetic moment.

To go beyond this description, it is necessary to start from the Dirac-equation or, if one wants to keep two-component wavefunctions, the Pauli-equation that can be derived from the Dirac-equation keeping terms up to second order in $c$, the velocity of light. (We consider here the single-particle equations, adopting the view that the Kohn-Sham equations are also in a single-particle form that can be extended by relativistic terms.) The Hamiltonian of the Pauli-equation can be written as (Bethe and Salpeter 1977):

$$H_{\text{Pauli}} = H_{\text{NR}} + H_{\text{SR}} + \frac{\hbar}{(2mc)^2}\underline{\boldsymbol{\sigma}} \cdot (\boldsymbol{\nabla} V(\mathbf{r}) \times \mathbf{p}) - \frac{e\hbar}{2mc}\underline{\boldsymbol{\sigma}} \cdot \mathbf{B}(\mathbf{r}).\qquad(11)$$

Here, $H_{\text{NR}}$ is the usual Schrödinger-type Hamiltonian (with the momentum operator $\mathbf{p}(\mathbf{r})$ extended by the vector potential $\frac{e}{c}\mathbf{A}(\mathbf{r})$) and $H_{\text{SR}}$ contains scalar-relativistic terms that do not contain spin-operators:

$$H_{\text{SR}} = -\frac{1}{2mc^2}(E + eV(\mathbf{r}))^2 + \frac{\hbar^2}{8m^2c^2}\nabla^2 V(\mathbf{r}).\qquad(12)$$

These are the mass-velocity correction containing the non-relativistic energy $E$ and the Darwin term, relevant mostly for $s$ electrons. The last two terms in Eq. 11 represent the spin-orbit interaction and the interaction of the spin with an external magnetic field. This last term was already included ad hoc in the potential matrix $\underline{v}(\mathbf{r})$. It may seem odd that in DFT calculations the magnetic field is normally considered via this B-field, while the mentioned vector potential contribution is neglected. Indeed, the latter is responsible for diamagnetic effects and has been discussed in the context of calculating NMR shielding tensors (Pickard and Mauri 2001), but requires a rather special treatment that goes beyond this introduction.

The physical interpretation of the spin-orbit coupling (SOC) term in Eq. (11) is most easily seen in an atom, where the gradient of the central potential can be written $\boldsymbol{\nabla} V(\mathbf{r}) = \frac{\partial V(r)}{\partial r}\frac{\mathbf{r}}{r}$. Then,

$$\frac{\hbar}{(2mc)^2}\underline{\boldsymbol{\sigma}} \cdot (\boldsymbol{\nabla} V(\mathbf{r}) \times \mathbf{p}) = \frac{\hbar}{(2mc)^2}\frac{1}{r}\frac{\partial V(r)}{\partial r}\underline{\boldsymbol{\sigma}} \cdot (\mathbf{r} \times \mathbf{p}) = \xi\underline{\boldsymbol{\sigma}} \cdot \mathbf{L},\qquad(13)$$

where $\xi$ is called the SOC constant and $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is the (angular) orbital momentum operator. In this form, the SOC term resembles a coupling between spin- and orbital momentum, although one has to keep in mind that it is rather the (orbital) motion of the electron in a potential gradient (i.e., an electric field that is Lorentz-transformed into a magnetic field) that couples to the spin. This also applies to a linear motion of an electron, where the SOC term manifests in the so-called Rashba-effect (Bychkov and Rashba 1984).
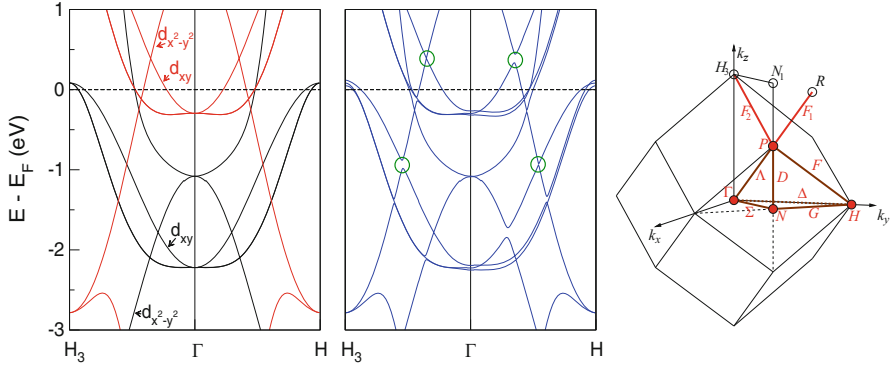
**Fig. 2** Band structure of bcc Fe without (left) and with (middle) SOC effects. The Brillouin zone is shown on the right. (From the Bilbao Crystallographic Server, http://www.cryst.ehu.es). Left: Majority and minority spin states are shown in black and red, respectively. The character of the orbitals forming the bands along $H_3 - \Gamma$ is indicated. With SOC the spin-character of the states is mixed and some band degeneracies are lifted. Crossings discussed in the text are marked with green circles. The magnetization is assumed to point in $y$ direction

Due to this coupling between the potential landscape (as, e.g., formed by the atomic lattice in a solid) and the electron spin, the orientation of the magnetization in the solid is no longer arbitrary but certain directions are preferred. This is one source of magnetic anisotropy, the so-called magnetocrystalline anisotropy, that determines the "easy axis" of magnetization. (Other sources, like the dipole-dipole interaction are not included in the single particle equation (11) and have to be added, e.g., from a classical calculation of the dipole sum.)

The effect of spin-orbit interaction in a magnetic system can be seen as a certain symmetry breaking that leads not only to a magnetic anisotropy, but also other effects like magnetostriction. This can be seen most clearly from a consideration of the band structure in two directions that are equivalent without SOC, but get non-equivalent when SOC is taken into account. As an example Fig. 2 shows the band structure of bulk bcc Fe along a high-symmetry direction in reciprocal space ($\Gamma - H$). In this direction, without SOC, several band crossings can be detected, e.g., between $d_{x^2-y^2}$ states and $d_{xy}$ states of same and different spin. The SOC term, Eq. (13), allows now hybridization between these states of the same spin to form an orbital moment in the direction of magnetization ($\Gamma - H$ corresponding to $y$) and these states of opposite spin in $\Gamma - H_3$ direction. Thus, gaps form at different positions in the formerly equivalent directions, which leads to a symmetry reduction from cubic to tetragonal in this case. These gap openings were recently also seen experimentally (Młyńczak et al. 2016).

Moreover, the spin-orbit terms gives rise to a finite orbital momentum that can be easily obtained in the vicinity of an atom $\nu$ as

$$\mathbf{M}_\nu^{\mathrm{orb}} = -\mu_B \sum_i \langle \phi_i | \mathbf{L} | \phi_i \rangle_\nu \tag{14}$$

where the integration is performed in a sphere around the atom. To obtain the orbital moment in an infinite periodic solid a more involved theory is necessary (Thonhauser et al. 2005). It should be noted that the orbital moments obtained by LSDA (and also GGA) calculations are typically too small, e.g., for Fe and Co values of 0.05 and 0.08 $\mu_B$ have been obtained (Beiden et al. 1998). Compared to experimental values (i.e., the difference between total and spin-momentum in Table 2) these values are reduced by almost 50%.

In absence of spin-flip terms (i.e., when the majority and minority band are well separated by the exchange interaction), SOC changes the total energy of a system in second-order perturbation theory as (van der Laan 1998):

$$\delta E = \sum_{i,j} \frac{\langle \phi_i | H_{so} | \phi_j \rangle \langle \phi_j | H_{so} | \phi_i \rangle}{\varepsilon_i - \varepsilon_j} f(\varepsilon_i) \left[ 1 - f(\varepsilon_j) \right] \approx -\frac{1}{4} \xi \hat{\mathbf{S}} \cdot \left[ \left\langle \mathbf{L}^\downarrow \right\rangle - \left\langle \mathbf{L}^\uparrow \right\rangle \right] \tag{15}$$

where $H_{so}$ is the SOC Hamiltonian, $f$ is the Fermi function giving the occupation of the state, $\hat{\mathbf{S}}$ is the direction of the spin moment, and $\mathbf{L}^{\downarrow(\uparrow)}$ is the orbital moment of the spin-down (up) bands. If the spin-up band is completely filled, we see that energy change, $\delta E$, is proportional to the size of the orbital moment. The magnetocrystalline anisotropy energy (MAE), i.e., the difference of $\delta E$ for two different magnetization directions, will be proportional to the difference in the orbital moments as first derived by Bruno (1989).

When calculating the MAE from DFT one has to keep in mind that the resulting energy differences are often very small, for cubic metals like Fe or Ni even in the $\mu$eV per unit cell range (see Table 3). Therefore, the numerical parameters, in particular those determining the sampling of reciprocal space, have to be extremely well converged (Trygg et al. 1995). To make the numerical effort tractable, often these calculations are not performed self-consistently, but SOC is added on top of a non- or scalar-relativistic calculation using the so-called force theorem (Weinert et al. 1985). Although the experimental values of the MAE for the cubic elements listed in Table 3 are hard to reproduce quantitatively, the easy axis is correctly reproduced in LSDA (Halilov et al. 1998).

**Table 3** Magnetic anisotropy energy (in $\mu$eV per atom) of the ferromagnetic elements obtained by DFT calculations with spin-orbit (SO) coupling included as obtained by Trygg et al. (1995) [1] and Halilov et al. (1998) [2]. In addition, the results including the orbital polarization correction (OP) are cited from [1]. The experimental data (exp.) are quoted from both references. For the fcc structures the anisotropy between [001] and [111] direction is given and positive values indicate that the latter magnetization direction is preferred. The values for Co (hcp) show the energy difference between the [0001] and [10$\bar{1}$0] direction

| Element | SO[1] | OP[1] | SO[2] | exp. |
|---------|-------|-------|-------|------|
| Fe (bcc) | –0.5 | –1.8 | –2.6 | –1.4 |
| Co (hcp) | –29 | –110 | | –65 |
| Co (fcc) | 0.5 | 2.2 | 2.4 | 1.3–1.8 |
| Ni (fcc) | –0.5 | –0.5 | 1.0 | 2.7 |

**Table 4** Magnetic anisotropy energy (in meV per formula unit) of the $L1_0$ phases of FePt and CoPt obtained in LSDA with spin-orbit (SO) coupling and orbital polarization (OP) by Ravindran et al. (2001) [1], Shick and Mryasov (2003) [2] and Kota and Sakuma (2014) [3], where perturbation theory (PT) and force theorem (FT) were used. Positive numbers indicate that the [001] direction is preferred (relative to [100]). The experimental data (exp.) are quoted from Ravindran et al. (2001)

| Element | SO[1] | OP[1] | SO[2] | PT[3] | FT[3] | Exp. |
|---------|-------|-------|-------|-------|-------|------|
| FePt    | 2.73  | 2.89  | 2.68  | 2.41  | 1.90  | 0.88 |
| CoPt    | 1.05  | 1.64  | 1.03  | 0.77  | 0.68  | 1.0–1.67 |
| FePd    | 0.15  | 0.34  |       | 0.33  | 0.29  | 0.48–0.63 |

As suggested by Eq. (15), the underestimation of the orbital moments in LSDA also leads to MAEs that are systematically too small. Techniques like the orbital polarization correction (OP) were introduced that increase both quantities leading to better agreement with experimental values (Trygg et al. 1995). Another way to improve results is to tale into account strong correlation effects via the Hubbard $U$ approximation (LSDA+$U$), as shown by Yang et al. (2001) or, for binary alloys by Shick and Mryasov (2003). For uniaxial systems, where the energies go up to the meV scale, the agreement is generally better as can be seen from Table 4.

Although in the examples listed above only the energies of two magnetization directions were compared, the MAE describes more generally the dependence of the energy on the spin orientation. In terms of the general two-spin interaction, $\underline{J}_{\nu\nu'}$, the magnetic anisotropy can be expressed as the traceless symmetric on-site part ($\nu = \nu'$) denoted as $\underline{K}_\nu$. It is a matrix with (at most) six coefficients that can be obtained by independent calculations.

One should keep in mind that a DFT calculation with SOC also gives rise to interactions described by the symmetric part of the $\underline{J}$ matrix (so-called pseudodipolar interactions) and the antisymmetric part of the $\underline{J}$ [Dzyaloshinskii-Moriya interaction (Dzialoshinskii 1957)] where the strength of the latter is first order in SOC, while the former arises in second order (Moriya 1960). Although small, these interactions can be extracted from DFT calculations as will be outlined below.

# 3 Ground State Calculations

## 3.1 Magnetic Order

It was shown that spin-polarized versions of DFT allow calculating also metastable magnetic (or non-magnetic) configurations and the result depends on the initial configuration. This makes determining the magnetic ground state of a specific system by DFT a rather challenging problem. Different strategies can be applied: like in molecular-dynamics calculations, (ab initio) spin-dynamics allows exploring the magnetic degrees of freedom and finding the global energy minimum that corresponds to the ground state configuration. A different possibility is to determine

the magnetic interactions between the atoms by DFT calculations which are then mapped onto a model (in the simplest case a classical Heisenberg model) that is solved analytically or numerically. In both cases a discretization of the (vector) magnetization density in terms of magnetic moments is introduced. Assume that at the positions of the magnetic atoms the intra-atomic exchange interaction is large and there are robust magnetic moments that can be assigned to each lattice site, e.g., within some sphere centered at the nucleus, $\nu$ (at a position $\mathbf{R}_\nu$). Then the magnetization density, $\mathbf{m}(\mathbf{r})$, can be approximated as

$$\mathbf{m}(\mathbf{r}) = M_\nu \hat{\mathbf{e}}_\nu, \tag{16}$$

where $M_\nu$ is the magnetization and $\hat{\mathbf{e}}_\nu$ is the magnetization direction at that site. A magnetic state is in this case characterized by a set of directions, $\{\hat{\mathbf{e}}_\nu\}$, of all the atoms in the magnetic unit cell. Of course there are the well-known ferromagnetic (FM) and antiferromagnetic (AFM) states, but already for the latter several possible unit cells come in mind. E.g., in an face-centered cubic (fcc) material, an AFM order can occur for ferromagnetically ordered planes in (001), (111), or (110) directions (type I, II or III AFM order, respectively) to name just the few possibilities that can be realized by a Heisenberg model with nearest-neighbor (nn) and next-nearest-neighbor (nnn) interactions.

On a simple level, one can "relax" the directions of the magnetization at the atoms like a relaxation of the atomic structure (e.g., at a surface) is done. In the DFT self-consistency cycle the output magnetizations $\{\hat{\mathbf{e}}_\nu^{\text{out}}\}$ in general deviate from the input values $\{\hat{\mathbf{e}}_\nu^{\text{in}}\}$. The magnetization directions change to minimize the total energy (cf. Fig. 3). The final magnetic state, that is reached, in general depends on the starting point of the calculation and a more elaborate technique is needed to avoid being trapped in some local minimum of $E[\{\hat{\mathbf{e}}_\nu\}]$.

To this end one can develop an equation of motion for the magnetization of an atom. To keep things simple it is again assumed that the magnetization stays collinear within the vicinity of the atom. Starting from the Hamiltonian of Eq. (9) and assuming that the external potential matrix, $\underline{v}(\mathbf{r})$, has been chosen to be diagonal and the exchange-correlation potential is separated into diagonal and off-diagonal parts, following Antropov et al. (1995, 1996) it is possible to set up a time-dependent analog of Eq. (9):

$$i\frac{d\Phi}{dt} = \left[ H_d - \underline{\boldsymbol{\sigma}} \cdot \mathbf{B}(\mathbf{r}, t) \right] \Phi \quad \text{where} \quad \Phi = \begin{pmatrix} \phi_\uparrow(\mathbf{r}, t) \\ \phi_\downarrow(\mathbf{r}, t) \end{pmatrix}, \tag{17}$$

and $H_d$ is the Hamiltonian that contains now only diagonal parts.

Separating the evolution of the magnetization into fast (value of the magnetization) and slow (direction of the magnetization) degrees of freedom the fast part are described quantum-mechanically, while the latter can be treated on a semiclassical level. At a given time, $t$, the time-independent version of Eq. (17) can be solved for a given magnetization characterized by $\{\hat{\mathbf{e}}_\nu\}$. An equation of motion for the
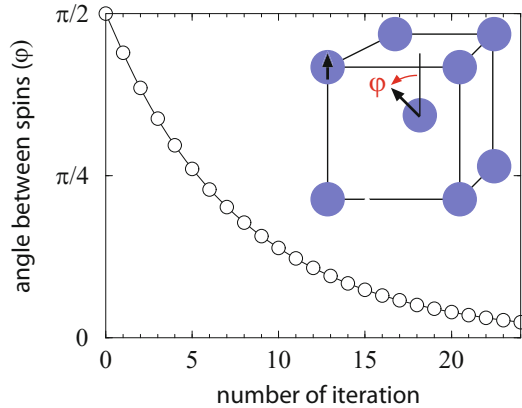
**Fig. 3** Relaxation of the spin direction of a Fe atom in bulk bcc Fe calculated in a simple cubic unit cell. The evolution of the angle $\varphi$ between the magnetic moments of two Fe atoms is shown as a function of the number of the self-consistency steps of the DFT calculation. Starting from a 90° canting of the spins, the FM ground state is reached after approx. 30 iterations. It is important to avoid high-symmetry states (e.g., AFM) as starting points. (Adapted with permission from Kurz et al. 2004. Copyrighted by the American Physical Society)

magnetization $\mathbf{m}(\mathbf{r}, t)$ can be obtained by multiplying Eq. (17) from the left with $-\mu_B \Phi^* \underline{\boldsymbol{\sigma}}$ and adding the complex conjugate equation. Comparing to the time derivative of the magnetization (cf. Eq. (5)), the equation of motion reads

$$\frac{d\mathbf{m}(\mathbf{r}, t)}{dt} = 2\mathbf{m} \times \mathbf{B} + \frac{i}{2}\nabla(\Phi^* \underline{\boldsymbol{\sigma}} \cdot \nabla\Phi - c.c.). \tag{18}$$

The second term on the right side is complicated and describes longitudinal changes of the magnetization, which will not be considered on this level. Omitting this term, Eq. (18) describes the precession of the magnetization direction at an atom under the influence of the magnetic field generated by the atom itself and other atoms of the crystal.

Equation (18) can be simplified using Eq. (16) and one can write for the evolution of the magnetization direction in atom $\nu$

$$\frac{d\hat{\mathbf{e}}_\nu}{dt} = -\frac{2}{\mu_B}\hat{\mathbf{e}}_\nu \times \mathbf{I}_\nu, \tag{19}$$

where $\mathbf{I}_\nu = \mu_B \mathbf{B}$. If the effect of other fields acting on a magnetization direction has to be taken into account explicitly, they can be included in Eq. (19). E.g., for contributions stemming from the spin-orbit interaction (magnetic anisotropy) or dipole-dipole interaction, these fields can be added in the form $\mathbf{I} = \mathbf{I}_\nu + \mathbf{I}_{SO} + \mathbf{I}_{d-d}$. More general expressions of Eq. (19), suitable for spin-dynamics with finite temperatures included, can be found in the paper of Antropov et al. (1996). It should

be noted that these equations are the starting point to describe many properties beyond the ground state on an ab initio basis e.g., shown by Skubic et al. (2008).

## 3.2    Magnetic Interactions

Although spin-dynamic simulations are a good way to explore magnetic ground states, they are computationally rather expensive and require additional analysis to get insight into the interactions leading to a specific magnetic order. For an interpretation of the results the system is usually mapped to a classical model spin Hamiltonian, e.g.,

$$H_{\mathrm{mod}} = - \sum_{\langle \nu\nu' \rangle} J_{\nu\nu'} \mathbf{M}_\nu \cdot \mathbf{M}_{\nu'} + \sum_\nu \mathbf{M}_\nu \underline{K}_\nu \mathbf{M}_\nu + \sum_\nu \mathbf{B} \cdot \mathbf{M}_\nu, \qquad (20)$$

where the first term on the right side captures the Heisenberg-like (scalar, two-spin) interactions and the sum runs over all pairs of atoms $\langle \nu, \nu' \rangle$, the second term is the magnetic anisotropy and the last term represents the interaction of the spins with an external $\mathbf{B}$ field (if needed). As mentioned above, the magnetic moments $\mathbf{M}_\nu$ are directly available from the DFT results and the anisotropy $\underline{K}_\nu$ can be obtained from relativistic total energy calculations.

To extract the $J_{\nu\nu'}$ coefficients in Eq. (20), it is in principle possible to compare the total energies of ferromagnetic and different antiferromagnetic states from non-relativistic DFT calculations and fit these differences to a classical spin Hamiltonian. As example in Fig. 4 the case of a square lattice is shown, as e.g., realized by an Fe monolayer on a (001) oriented bcc substrate. Ferriani et al. (2007) studied this system for Ta and W as substrate and found a transition from a FM ground state for Ta to a $c(2 \times 2)$ AFM state for Fe/W(001). In a nearest-neighbor (nn) Heisenberg model this transition corresponds to a sign change of the nn coupling constant $J_1$. Including also next-nearest-neighbor (nnn) interactions, this model also allows (for negative $J_2$) for another ground state, a row-wise AFM state with a $p(2 \times 1)$ unit cell
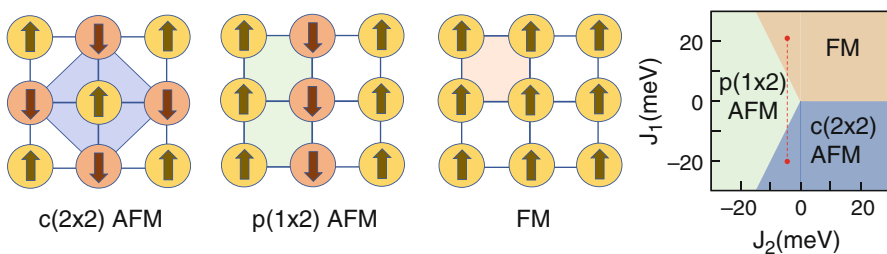


**Fig. 4** Two antiferromagnetic (AFM) and the ferromagnetic (FM) structure on a square lattice. The magnetic unit cells are indicated by the shaded areas. The phase diagram for the classical Heisenberg model with nearest-neighbor (nn) and next-nearest-neighbor (nnn) interaction is shown on the right. The red dashed line indicates a phase transition as discussed in the text. (Adapted with permission from Ferriani et al. 2007. Copyrighted by the American Physical Society)
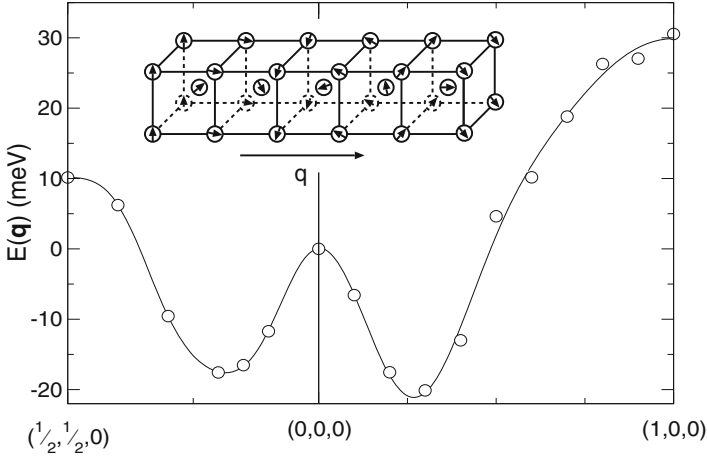
**Fig. 5** Total energy of bcc-Eu as a function of the spin-spiral **q**-vector (in units of $\frac{2\pi}{a}$ where $a$ is the lattice constant) in (110) and (100) direction calculated in the LDA+$U$ method. The symbols represent calculated values, the lines are a fit of $E(q) = \sum_{n=0}^{5} c_n \cos(n\pi q)$. The energy minimum at **q** $= (1.63, 0, 0)/a$ (also visualized in the inset) corresponds quite well to the experimental ground state (Turek et al. 2003b)

as shown in Fig. 4. By calculating these three magnetic states for alloyed substrates, $Ta_x W_{1-x}$, one can see that from $x = 1$ to $x = 0$ $J_1$ changes from about $+20$ to $-20$ meV while $J_2$ remains rather constant at $-5$ meV. In the phase diagram in Fig. 4 this corresponds to a change along the red dashed line and one can see that for $x = 0.5$ a $p(2 \times 1)$ AFM structure can be expected as magnetic ground state (Ferriani et al. 2007) (Fig. 5).

Of course, more magnetic structures can be added to the analysis and further interactions, $J_i$ for $i > 2$, can be extracted. A systematic way to obtain these $J$'s is given by the calculation of so-called spiral spin-density waves (SSDWs), general solutions of the classical Heisenberg model. The magnetic moments in a spiral with propagation vector **q** are given as:

$$\mathbf{M}_\nu = M \left( \hat{\mathbf{e}}_x \cos(\mathbf{q} \cdot \mathbf{R}_\nu) + \hat{\mathbf{e}}_y \sin(\mathbf{q} \cdot \mathbf{R}_\nu) \right), \tag{21}$$

where the unit vectors $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$ just have to be perpendicular to each other, otherwise their directions are (in absence of SOC) arbitrary. For simplicity it is assumed that the size of the magnetic moment, $M$, is not site dependent. The wave-length of the spin-spiral is given by $\lambda = \frac{2\pi}{q}$ and it seems that huge super-cells are necessary to capture these SSDWs. But using the generalized Bloch theorem (Herring 1966; Sandratskii 1991) these magnetic structures can be conveniently calculated in the chemical unit cell as long as no spin-orbit coupling is considered. In this way the total energy $E(\mathbf{q})$ can be calculated on a grid of **q** vectors and, assuming that $M$ remains constant, directly related to $J(\mathbf{q})$ which is the Fourier transform of the Heisenberg interaction constants in real space. By

Fourier transformation the constants $J_{\nu\nu'}$ can thus be recovered. This procedure can be generalized to systems with different atomic species in the unit cell, where intra-sublattice $J$'s and inter-sublattice $J$'s have to be considered (Ležaić et al. 2013).

Spin-spiral structures, as given by Eq. (21), also occur as ground states in many magnetic systems. A well-investigated system is bcc Eu, a member of the lanthanides, where the spins are rotated between neighboring (001) planes by 47° (Turek et al. 2003b). DFT calculations of SSDWs with $\mathbf{q}$ along the $(1, 0, 0)$ direction indeed show a minimum of $E(\mathbf{q})$ at the experimentally observed position. Also along the $(1, 1, 0)$ direction a local minimum is found and experiments indeed report that under an external magnetic field such state can be realized (Millhouse and McEwen 1973). It should be further mentioned that in these calculations the so-called LDA+$U$ method (Anisimov et al. 1997) was used to capture the correlated nature of the $4f$ electrons in Eu.

As mentioned above, SSDWs are the general solutions of the classical Heisenberg model (Yosida 1996). Therefore, if a system described by this model, a procedure to find the magnetic ground state was presented. However, for a system with spins larger than $1/2$, higher order interactions are allowed that involve more than two spins, e.g., the biquadratic interaction $B_{\nu\nu'}(\mathbf{S}_\nu \cdot \mathbf{S}_{\nu'})^2$ or other four-spin interactions. These terms can couple two or more spin-spirals and possibly lead to solutions with lower energy than all SSDWs. E.g., the $p(2 \times 1)$ structure in Fig. 6 can be regarded as spin spiral with $\mathbf{q}_1 = \left(0, \frac{1}{2}\right)$ and a 90° rotated SSDW with $\mathbf{q}_2 = \left(\frac{1}{2}, 0\right)$ is degenerate with this structure. A linear combination $\frac{1}{2}(\mathbf{q}_1 + \mathbf{q}_2)$, where the spins of neighboring atoms are 90° rotated to each other (see Fig. 6), is energetically degenerate to the constituting spin spirals in the classical Heisenberg model, but a Hamiltonian that includes the biquadratic interaction can differentiate between the former double-$\mathbf{q}$ and the latter single-$\mathbf{q}$ structures. In a DFT calculation it is found that indeed the double-$\mathbf{q}$ structure has lower total energy (Ferriani et al. 2007). This is a sign for the importance of higher-order interactions that is not unexpected in a system with large spin moment.



p(2x1) AFM (0meV)    2q (−3meV)    p(1x2) AFM (0meV)
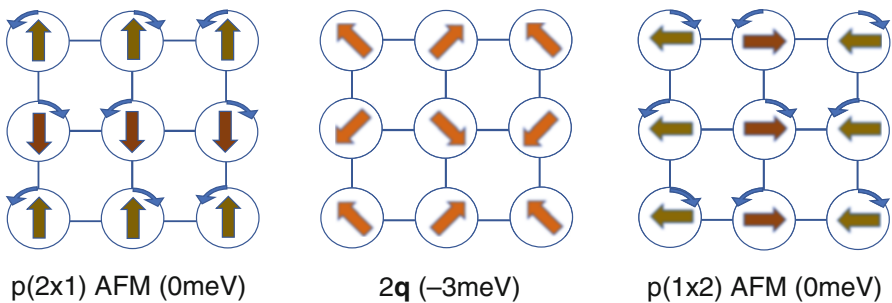
**Fig. 6** The 2$\mathbf{q}$ structure shown in the middle is a superposition of the $p(2 \times 1)$ and $p(1 \times 2)$ structure. Alternatively, it can be constructed by rotation of the spins in the collinear structures by 45° as indicated by the blue arrows. For Fe on a Ta/W-alloy substrate, the 2$\mathbf{q}$ structure is 3 meV lower in energy than the two shown antiferromagnetic structures (Ferriani et al. 2007)

All SSDWs and the AFM structures shown in Fig. 4 [that are actually spin spirals with $\mathbf{q}_{c(2\times2)} = \left(\frac{1}{2}, \frac{1}{2}\right)$, $\mathbf{q}_{p(1\times2)} = \left(\frac{1}{2}, 0\right)$ and $\mathbf{q}_{FM} = (0, 0)$] are special in the sense that in these structures there is no torque on the magnetic moments due to symmetry. I.e., in the DFT self-consistency cycle the spin direction does not change in the sense described in Sect. 3.1. This is different for the spin-structures that connect the states shown in Fig. 6. To calculate the total energy of such spin structures correctly, constrained DFT calculations (Dederichs et al. 1984) have to be performed, i.e., a transversal $\mathbf{B}_\perp$ field has to be applied that keeps the local moment in a prescribed direction (Kurz et al. 2004). This constrained DFT scheme can also be used to study longitudinal fluctuations of the magnetic moments, $E(|\mathbf{M}|)$. These so-called fixed spin-moment calculations can be very useful to explore magnetic phases, in particular when multiple solutions (high-spin and low-spin phases) are present (Moruzzi et al. 1986).

It was mentioned above that SSDW calculations using the generalized Bloch theorem can only be performed in the absence of SOC. Nevertheless, it is possible to include SOC in first order perturbation theory to the spin-spiral calculations (Heide et al. 2009). In this way, it is possible to access the Dzyaloshinskii-Moriya interaction (DMI) mentioned at the end of Sect. 2.2 in a rather systematic way. This relativistic interaction can then be added to the model Hamiltonian (20) to describe large-scale spin structures. For example the DMI can stabilize a SSDW with unique rotational sense (Bode et al. 2007) or it leads to chiral magnetic skyrmions in complex superlattices with several nanometers of diameter (Dupé et al. 2016). It can also be used as input for micromagnetic simulations of even larger systems (see Schweflinghaus et al. (2016) for an relation of atomistic and micromagnetic parameters).

# 4 Beyond the Ground State

Although a more detailed account of spin excitations and finite temperature magnetism will follow in the next chapters, this section is intended to show the link between non-collinear DFT calculations, as described above, and the modeling of properties beyond the magnetic ground state. Moreover, the application range and limitations of some simple models are indicated.

## 4.1 Magnetic Fluctuations

Up to now, flat spin-spirals as defined in Eq. (21) were discussed, i.e., the spins were assumed to be rotating in the plane defined by $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$. These SSDWs can be generalized to conical spin-spirals as

$$\mathbf{M}_\nu = M \left( \hat{\mathbf{e}}_x \cos(\mathbf{q} \cdot \mathbf{R}_\nu) \sin(\vartheta) + \hat{\mathbf{e}}_y \sin(\mathbf{q} \cdot \mathbf{R}_\nu) \sin(\vartheta) + \hat{\mathbf{e}}_z \cos(\vartheta) \right), \qquad (22)$$

where $2\vartheta$ defines the opening angle of the cone. Magnetic structures described by Eq. (22) occur as magnetic ground states, e.g., in $LaMn_2Ge_2$ (Di Napoli et al. 2004), or in strong magnetic fields as described by the Hamiltonian (20).

Since conical spin-spirals also resemble snap shots of a single magnon in a ferromagnet at a fixed time, the SSDWs introduced above are sometimes called "frozen magnons". Calculations for different (small) $\mathbf{q}$ values (e.g., using the generalized Bloch theorem) can be used to simulate the effect of temperature on a magnetic system in the adiabatic approximation. At very low temperatures, when magnons with long wavelengths dominate and no Stoner excitations are present, this is a reasonable approximation. In the long wavelength limit, i.e., around $\mathbf{q} = 0$, the spin-wave dispersion behaves almost quadratically and can be described as $E(\mathbf{q}) = Dq^2$. The spin stiffness, $D$, characterizes the magnetic properties of a ferromagnet at low temperatures and can be calculated from SSDW calculations directly or from the exchange coupling constants in real space. Consider e.g., the Heisenberg term of Eq. (20) in one dimension: Then

$$\delta E(\mathbf{q}) = E(\mathbf{0}) - E(\mathbf{q}) = +2M^2 \sum_{\nu>0} J_{0\nu} \left(1 - \cos(\mathbf{q}\mathbf{R}_\nu)\right) \approx 2M^2 q^2 \sum_{\nu>0} J_{0\nu} \mathbf{R}_\nu^2 \tag{23}$$

and $D$ is obtained as $2M^2 \sum_{\nu>0} J_{0\nu} \mathbf{R}_\nu^2$. In more dimensions $D$ is a tensorial quantity and the exact expression depends on the crystal symmetry. Some results of an isotropic $D$ obtained from DFT calculations for cubic systems are given in Table 5. For Fe and Co agreement with experimental data is reasonable, but for Ni most methods fail to reproduce the experimental spin stiffness. More about spin-waves can be found e.g., in the review of Staunton (1994) and in a following chapter. Calculations of the full magnon spectra of Fe and Ni can be found e.g., in the work of Halilov et al. (1997). The agreement with experimental spectra, even for larger $\mathbf{q}$ values is quite satisfactorily, therefore these calculations can also serve as a starting point to calculate finite temperature properties, like the Curie temperature. One has to keep in mind, however, that for other systems it is not guaranteed that they can be treated in such simple model.

In the discussion of spin waves so far the influence of spin-orbit coupling was neglected. Therefore, the so obtained "spin-excitations" are always gapless while in a system with magnetic anisotropy there is an energy to overcome for a finite

**Table 5** Calculated and experimental spin-wave stiffness ($D$) for Fe, Co and Ni. The theoretical values were obtained in different approximations as described by Rosengaard and Johansson (1997) [th.(1)], Kübler (2000) [th.(2)], Shallcross and coworkers (2005) [th.(3)] and Pajda et al. (2001) [th.(4)], experimental data were taken as cited in these references

| $D$ (meV Å$^2$) | th.(1) | th.(2) | th.(3) | th.(4) | Exp. |
|---|---|---|---|---|---|
| Fe (bcc) | 247 | 355 | 322,313 | 250 | 280,314,330 |
| Co (fcc) | 502 | 510 | 480,520 | 663 | 510,580 |
| Ni (fcc) | 739 | 790 | 541,1796 | 756 | 422,550,555 |

**q** value. Moreover, the Dzyaloshinskii-Moriya interaction leads to a finite energy difference between clockwise and anticlockwise rotating magnons (Udvardi and Szunyogh 2009).

## 4.2 Ordering Temperatures

Extending the description of finite-temperature magnetic properties of bulk ferromagnets to the point where transverse fluctuations finally destroy the magnetic order, here the simplest model is presented, the so-called mean field approximation (MFA). A selected site 0 experiences an effective field from all other atoms; At $T = 0$ it is proportional to $M \sum_{\nu \neq 0} J_{0\nu} = M J_0$, while at $T > 0$ this field is reduced by thermal fluctuation on the sites $\nu$:

$$B_{\text{eff}} = \sum_\nu J_{0\nu} \langle M(\mathbf{R}_\nu) \rangle, \tag{24}$$

where the thermal average of the projection of the magnetization at site $\nu$ on the magnetization at site 0 is denoted as $\langle M(\mathbf{R}_\nu) \rangle$. In this model the temperature where the average magnetization vanishes (for ferromagnetic systems the Curie temperature) is given by

$$k_B T_C^{\text{MFA}} = \frac{2}{3} J_0 M^2 \eta, \tag{25}$$

where $\eta \to 1$ in the classical limit (Liechtenstein et al. 1986). In many cases the MFA overestimates $T_C$ and other models, like the random phase approximation (RPA) are preferable. Moreover, in lower-dimensional systems the MFA cannot be applied, e.g., a finite the magnetic anisotropy is essential to stabilize magnetic order in two dimensions (Bander and Mills 1988). However relativistic DFT calculations provide the necessary ingredients to calculate also $T_C$ in two dimensional systems (Udvardi and Szunyogh 2009). Also for antiferromagnets (or, generally spin-spiral states) expressions for the ordering temperature, the Néel temperature $T_N$, can be derived (Turek et al. 2003b). There exist several more methods to calculate critical temperatures from DFT results, e.g., the Monte Carlo technique (MC) allows studying finite-temperature magnetic properties by implementation of a Heisenberg Hamiltonian, possibly with extensions like the magnetic anisotropy like in Eq. (20). Some applications for multi-sublattice systems are shown in the work of Ležaić et al. (2013).

The Curie temperature of Fe, Co, Ni and Gd obtained in different approximations with parameters from DFT calculations are presented in Table 6. Compared to RPA, the MFA overestimates $T_C$ by 25–50% in these cases. For Fe and Co RPA gives quite good estimates of the Curie temperature, while for Ni $T_C$ is underestimated in both approximations. From Table 1 it can be seen that this behavior is not universal and for some materials the MFA results compare very well with experimental data.

**Table 6** Calculated and experimental Curie temperature $T_C$ for some ferromagnetic materials. MFA[1] and RPA data for Fe, Co and Ni taken from Pajda et al. (2001), MFA[2] results and experimental values as quoted by Shallcross and coworkers (2005), while the MC results were obtained by Rosengaard and Johansson (1997). Spin dynamics (SD) calculations have been performed by Antropov (2005). Data for Gd can be found in the papers of Kurz et al. (2002) and Turek and coworkers (2003a)

| $T_C$ (K) | MFA[1] | MFA[2] | RPA | MC | SD | Exp. |
|-----------|--------|--------|-----|-----|-----|------|
| Fe (bcc) | 1414 | 550,1190 | 950 | 1060 | 1070 | 1044 – 1045 |
| Co (fcc) | 1645 | 1120,1350 | 1311 | 1080 | | 1388 – 1393 |
| Ni (fcc) | 397 | 320,820 | 350 | 510 | 470 | 624 – 631 |
| Gd (hcp) | 334 | | | | | 293 |

MC simulations work better for Ni and Fe, but give a too low $T_C$ for Co. Finally, the results of spin dynamics calculations, performed along the line sketched in Sect. 3.1, give results comparable to MC calculations for Fe and Ni, but have the advantage that they do not rely on a model Hamiltonian (Antropov 2005). Also in the work of Skubic et al. (2008) it can be seen that SD and MC results for bcc Fe are very similar.

# 5    Conclusion

Within this small chapter I tried to give an impression of the possibilities that DFT calculations can give to calculate magnetic properties of elemental magnets and some compounds, mainly focusing on bulk materials. I left out many classes of magnetic materials, e.g., transition metal oxides, where correlation effects complicate the picture and extensions like LDA + $U$ (Anisimov et al. 1997) or LDA + DMFT (Held et al. 2002) are required to capture many aspects of their properties. Of course, there are many more magnetic properties where DFT can be very useful for modeling, e.g., the coupling of electrons with the nuclear spin that has been neglected here. Nevertheless, I hope that this chapter gives at least a staring point to explore the power of ab initio methods to model magnetic systems.

# References

Anisimov VI, Aryasetiawan F, Lichtenstein AI (1997) First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LSDA+$U$ method. J Phys Condens Matter 9:767–808

Antropov V (2005) Magnetic short-range order above the Curie temperature of Fe and Ni. Phys Rev B 72:140406

Antropov VP, Katsnelson MI, van Schilfgaarde M, Harmon BN (1995) Ab initio spin dynamics in magnets. Phys Rev Lett 75:729–732

Antropov VP, Katsnelson MI, Harmon BN, van Schilfgaarde M, Kusnezov D (1996) Spin dynamics in magnets: equation of motion and finite temperature effects. Phys Rev B 54: 1019–1035

Bander M, Mills DL (1988) Ferromagnetism of ultrathin films. Phys Rev B 38:12015–12018

Beiden SV, Temmerman WM, Szotek Z, Gehring GA, Stocks GM, Wang Y, Nicholson DMC, Shelton WA, Ebert H (1998) Real-space approach to the calculation of magnetocrystalline anisotropy in metals. Phys Rev B 57:14247–14253

Bethe HA, Salpeter EE (1977) Quantum mechanics of one- and two-electron systems. Plenum, New York

Bode M, Heide M, von Bergmann K, Ferriani P, Heinze S, Bihlmayer G, Kubetzka A, Pietzsch O, Blügel S, Wiesendanger R (2007) Chiral magnetic order at surfaces driven by inversion asymmetry. Nature 447:190

Bruno P (1989) Tight-binding approach to the orbital magnetic moment and magnetocrystalline anisotropy of transition-metal monolayers. Phys Rev B 39:865–868

Bychkov YA, Rashba EI (1984) Properties of a 2D electron gas with lifted spectral degeneracy. JETP Lett 39:78–81

Dederichs PH, Blügel S, Zeller R, Akai H (1984) Ground states of constrained systems: application to cerium impurities. Phys Rev Lett 53:2512–2515

Di Napoli S, Llois AM, Bihlmayer G, Blügel S, Alouani M, Dreyssé H (2004) Magnetic structure and transport properties of noncollinear $LaMn_2X_2$ ($X$ = Ge, Si) systems. Phys Rev B 70:174418

Dirac PAM (1930) Note on exchange phenomena in the Thomas–Fermi atom. Proc Camb Philos Soc 26:376–385

Dupé B, Bihlmayer G, Böttcher M, Blügel S, Heinze S (2016) Engineering skyrmions in transition-metal multilayers for spintronics. Nat Commun 7:11779

Dzialoshinskii IE (1957) Thermodynamic theory of "weak" ferromagnetism in antiferromagnetic substances. Sov Phys JETP 5:1259

Eich FG, Gross EKU (2013) Transverse spin-gradient functional for noncollinear spin-density-functional theory. Phys Rev Lett 111:156401/1–5

Ferriani P, Turek I, Heinze S, Bihlmayer G, Blügel S (2007) Magnetic phase control in monolayer films by substrate tuning. Phys Rev Lett 99:187203

Friesner RA (2005) Ab initio quantum chemistry: methodology and applications. PNAS 102: 6648–6653

Galler A, Taranto C, Wallerberger M, Kaltak M, Kresse G, Sangiovanni G, Toschi A, Held K (2015) Screened moments and absence of ferromagnetism in FeAl. Phys Rev B 92: 205132

Halilov SV, Perlov AY, Oppeneer PM, Eschrig H (1997) Magnon spectrum and related finite-temperature magnetic properties: a first-principle approach. Europhys Lett 39:91

Halilov SV, Perlov AY, Oppeneer PM, Yaresko AN, Antonov VN (1998) Magnetocrystalline anisotropy energy in cubic Fe, Co, and Ni: applicability of local-spin-density theory reexamined. Phys Rev B 57:9557–9560

Heide M, Bihlmayer G, Blügel S (2009) Describing Dzyaloshinskii-Moriya spirals from first principles. Physica B 404:2678

Held K, Anisimov VI, Nekrasov IA, Eyert V, Keller G, Pruschke T, Vollhardt D, Blümer N, McMahan AK, Scalettar RT (2002) The LDA+DMFT approach to materials with strong electronic correlations. In: Grotendorst J, Marx D, Muramatsu A (eds) Quantum simulations of complex many-body systems: from theory to algorithms, Research Center Jülich. NIC series, vol 10. John von Neumann Institute for Computing, Jülich, pp 175–209

Herring C (1966) Exchange interactions among intinerant electrons. In: Rado GT, Suhl H (eds) Magnetism, vol IV. Academic, New York/London

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871

Ibañez Azpiroz J, dos Santos Dias M, Blügel S, Lounis S (2016) Zero-point spin-fluctuations of single adatoms. Nano Lett 16:4305–4311

Jones RO (2015) Density functional theory: its origins, rise to prominence, and future. Rev Mod Phys 87:897–923

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138

Kokalj A (2003) Computer graphics and graphical user interfaces as tools in simulations of matter at the atomic scale. Comput Mater Sci 28:155–168

Koller D, Tran F, Blaha P (2011) Merits and limits of the modified Becke-Johnson exchange potential. Phys Rev B 83:195134/1–10

Kota Y, Sakuma A (2014) Mechanism of uniaxial magnetocrystalline anisotropy in transition metal alloys. J Phys Soc Jpn 83(3):034715

Kübler J (2000) Theory of itinerant electron magnetism. International series of monographs in physics, vol 106. Clarendon Press, Oxford

Kurz P, Bihlmayer G, Blügel (2002) Magnetism and electronic structure of hcp Gd and the Gd(0001) surface. J Phys Condens Matter 14:6353–6371

Kurz P, Förster F, Nordström L, Bihlmayer G, Blügel S (2004) Ab initio treatment of non-collinear magnets with the full-potential linearized augmented planewave method. Phys Rev B 69:024415/1–15

Ležaić M, Mavropoulos P, Bihlmayer G, Blügel S (2013) Exchange interactions and local-moment fluctuation corrections in ferromagnets at finite temperatures based on noncollinear density-functional calculations. Phys Rev B 88:134403

Liechtenstein AI, Katsnelson MI, Antropov VP, Gubanov VA (1986) Local spin density functional approach to the theory of exchange interactions in ferromagnetic metals and alloys. J Magn Magn Mater 67:65–74

Millhouse AH, McEwen KA (1973) Neutron diffraction study of single crystal europium in an applied magnetic field. Solid State Commun 13:339–345

Młyńczak E, Eschbach M, Borek S, Minár J, Braun J, Aguilera I, Bihlmayer G, Döring S, Gehlmann M, Gospodarič P, Suga S, Plucinski L, Blügel S, Ebert H, Schneider CM (2016) Fermi surface manipulation by external magnetic field demonstrated for a prototypical ferromagnet. Phys Rev X 6:041048

Moriya T (1960) Anisotropic superexchange interaction and weak ferromagnetism. Phys Rev 120:91–98

Moruzzi VL, Janak JF, Williams AR (1978) Calculated electronic properties of metals. Pergamon, New York

Moruzzi VL, Marcus PM, Schwarz K, Mohn P (1986) Ferromagnetic phases of bcc and fcc Fe, Co, and Ni. Phys Rev B 34:1784–1791

Pajda M, Kudrnovský J, Turek I, Drchal V, Bruno P (2001) Ab initio calculations of exchange interactions, spin-wave stiffness constants, and Curie temperatures of Fe, Co, and Ni. Phys Rev B 64:174402

Pickard CJ, Mauri F (2001) All-electron magnetic response with pseudopotentials: NMR chemical shifts. Phys Rev B 63:245101/1–13

Ravindran P, Kjekshus A, Fjellvåg H, James P, Nordström L, Johansson B, Eriksson O (2001) Large magnetocrystalline anisotropy in bilayer transition metal phases from first-principles full-potential calculations. Phys Rev B 63:144409

Rosengaard NM, Johansson B (1997) Finite-temperature study of itinerant ferromagnetism in Fe, Co, and Ni. Phys Rev B 55:14975–14986

Sandratskii LM (1991) Symmetry analysis of electronic states for crystals with spiral magnetic order. I. General properties. J Phys Condens Matter 3:8565

Schweflinghaus B, Zimmermann B, Heide M, Bihlmayer G, Blügel S (2016) Role of Dzyaloshinskii-Moriya interaction for magnetism in transition-metal chains at Pt step edges. Phys Rev B 94:024403

Shallcross S, Kissavos AE, Meded V, Ruban AV (2005) An ab initio effective Hamiltonian for magnetism including longitudinal spin fluctuations. Phys Rev B 72:104437/1–8

Shick AB, Mryasov ON (2003) Coulomb correlations and magnetic anisotropy in ordered $L1_0$ CoPt and FePt alloys. Phys Rev B 67:172407

Singh DJ, Ashkenazi J (1992) Magnetism with generalized-gradient-approximation density functionals. Phys Rev B 46:11570–11577

Skubic B, Hellsvik J, Nordström L, Eriksson O (2008) A method for atomistic spin dynamics simulations: implementation and examples. J Phys Condens Matter 20(31):315203

Slater JC (1951) A simplification of the Hartree-Fock method. Phys Rev 81:385–390

Staunton JB (1994) Electronic structure of magnetic transition metallic materials. Rep Prog Phys 57:1289–1344

Thoene J, Chadov S, Fecher G, Felser C, Kübler J (2009) Exchange energies, Curie temperatures and magnons in Heusler compounds. J Phys D Appl Phys 43:084013

Thonhauser T, Ceresoli D, Vanderbilt D, Resta R (2005) Orbital magnetization in periodic insulators. Phys Rev Lett 95:137205/1–4

Trygg J, Johansson B, Eriksson O, Wills JM (1995) Total energy calculation of the magnetocrystalline anisotropy energy in the ferromagnetic 3$d$ metals. Phys Rev Lett 75:2871–2874

Turek I, Kudrnovský J, Bihlmayer G, Blügel S (2003a) Ab initio theory of exchange interactions and the Curie temperature of bulk Gd. J Phys Condens Matter 15(17):2771

Turek I, Kudrnovský J, Diviš M, Franek P, Bihlmayer G, Blügel S (2003b) First-principles study of the electronic structure and exchange interactions in bcc europium. Phys Rev B 68:224431

Udvardi L, Szunyogh L (2009) Chiral asymmetry of the spin-wave spectra in ultrathin magnetic films. Phys Rev Lett 102:207204

van der Laan G (1998) Microscopic origin of magnetocrystalline anisotropy in transition metal thin films. J Phys Condens Matter 10(14):3239

von Barth U, Hedin L (1972) A local exchange-correlation potential for the spin polarized case: I. J Phys C Solid State Phys 5:1629–1642

Weinert M, Watson RE, Davenport JW (1985) Total-energy differences and eigenvalue sums. Phys Rev B 32:2115–2119

White HW, Beaudry BJ, Burgardt P, Legvold S, Harmon BN (1975) Magnetic moments of ferromagnetic gadolinium alloys. AIP Conf Proc 29:329

Yang I, Savrasov SY, Kotliar G (2001) Importance of correlation effects on magnetic anisotropy in Fe and Ni. Phys Rev Lett 87:216405/1–4

Yosida K (1996) Theory of magnetism. Springer, Berlin/Heidelberg

# Many-Body Spin Excitations in Ferromagnets from First Principles

Christoph Friedrich, Mathias C. T. D. Müller, and Stefan Blügel

## Contents

**Abstract**

Electronic spin excitations are low-energy excitations that influence the properties of magnetic materials substantially. Two types of spin excitations can be identified, single-particle Stoner excitations and collective spin-wave excitations. They can be treated on the same footing within many-body perturbation theory. In this theory, the collective spin excitations arise from the correlated motion of electron-hole pairs with opposite spins. We present the theory in detail and discuss several aspects of an implementation within the full-potential linearized augmented plane-wave method. The pair propagation is described by the transverse magnetic susceptibility, which we calculate from first principles employing the ladder approximation for the $T$ matrix. The four-point $T$ matrix is represented in a basis of Wannier functions. By using an auxiliary Wannier set with suitable Bloch character, the magnetic response function can be evaluated for arbitrary $\mathbf{k}$ points, allowing fine details of the spin-wave spectra to be studied. The energy of the acoustic spin-wave branch should vanish in the limit $\mathbf{k} \to \mathbf{0}$,

C. Friedrich (✉) · M. C. T. D. Müller · S. Blügel
Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, Jülich, Germany
e-mail: c.friedrich@fz-juelich.de; s.bluegel@fz-juelich.de

which is a manifestation of the Goldstone theorem. However, this condition is often violated in the calculated acoustic magnon dispersion, which can partly be traced back to the choice of the Green function. In fact, the numerical gap error is considerably reduced when a renormalized Green function is used. As an alternative simple correction scheme, we suggest an adjustment of the Kohn-Sham exchange splitting. We present spin excitation spectra for the elementary ferromagnets Fe, Co, and Ni as illustrative examples and compare to model calculations of the homogeneous electron gas.

# 1 Introduction

Electronic spin excitations span a large range of energies, from high-energy single-particle Stoner excitations to low-energy collective spin-wave excitations. Therefore, they are present at all temperatures and play an important role for the physical properties of magnetic materials. For example, the specific heat (Doniach and Engelsberg 1966), the macroscopic magnetization (Bloch 1930), and the magnetic susceptibility (Moriya 1985) exhibit a characteristic temperature dependence which can be attributed to the low-energy spin waves (magnons) with excitation energies ranging from a few meV up to a few hundreds meV. In low-dimensional magnets, spin-wave excitations can even destroy the long-range magnetic order completely. This happens in the absence of magnetic anisotropy at any finite temperature (Mermin and Wagner 1966). As the temperature increases, in addition to collective magnon modes, single-particle spin-flip processes, so-called Stoner excitations, become increasingly important. They further contribute to the temperature variation of the magnetization and give rise to a damping of the magnon states.

Spin excitations play a role in many fields of fundamental and technological interest. They can contribute to the scattering of a propagating electron or hole in a magnetic material, leading to a renormalization of the quasiparticle band dispersion (Hofmann et al. 2009; Schäfer et al. 2004) and reducing the inelastic mean free path of hot electrons (Hong and Mills 1999, 2000). In magnetic recording applications, the creation of spin waves that accompanies each switching process in the storage medium sets physical limits on data rates and areal recording densities. A strong damping helps in dissipating the energy contained in the spin waves. The spin wave bus, on the other hand, utilizes spin waves as a means for information transmission between distant nanoscale devices (Khitun and Wang 2005). The damping through the creation of Stoner excitations is an undesirable effect in this case, limiting the distance over which information can be transmitted. The power consumption of such a spin wave bus is expected to be considerably lower than in charge-based devices. Finally, it is believed that the electron-electron interaction can become attractive through the exchange of magnons, which is a possible mechanism for the creation of Cooper pairs in high-temperature superconductors (Dagotto 1994; Scalapino 1995).

In quantum mechanics, the spin excited states are eigenstates of the many-body Hamiltonian with a net spin flip with respect to the many-body ground state. The magnetic response function or dynamical spin susceptibility, defined as a two-particle Green function in many-body perturbation theory, exhibits resonances at the corresponding eigenenergies. In an infinite system, these resonances – or poles – are, in general, not discrete but form a continuous distribution when plotted over the eigenenergy. This *spectral function* is given mathematically by the imaginary part of the magnetic response function and can be measured in inelastic neutron scattering experiments (Lowde et al. 1983) where the circularly polarized magnetic field of an incoming neutron beam disturbs the local magnetization density of the sample material. The magnetic response function is thus a central quantity in the theoretical study of magnetic materials.

For a theoretical description of spin dynamics, various formalisms have been established. Most theoretical studies to date employ the Heisenberg model, which relies on a separation of the magnetic degrees of freedom and the fast motion of the electrons. This model is defined in terms of exchange parameters, which can be obtained, for example, from constrained density functional theory (Rosengaard and Johansson 1997; Kübler 2009; Halilov et al. 1997). Spin waves with long wavelengths can be calculated efficiently. Single-particle Stoner excitations are neglected, though, and the linewidths of the spin-wave resonances, which are inversely proportional to the magnon lifetimes, are inaccessible. Furthermore, the Heisenberg model is strictly justified only for systems with localized moments, such as systems with rare-earth magnetic ions, but not for materials which are magnetized by the exchange-driven polarization of the spins of itinerant electrons. While the Heisenberg model still yields reasonable results for long-wavelength excitations in itinerant-electron magnets, results for short-wavelength excitations are unsatisfactory. For example, the multiple branches or gaps in the magnon dispersion of $3d$ ferromagnets cannot be captured (Cooke 1976).

Many-body perturbation theory (MBPT) provides a more general theoretical framework that works for systems with localized moments and for metallic magnets alike. Single-particle Stoner and collective spin excitations appear simultaneously as poles in the transverse magnetic susceptibility, which can be interpreted as describing the correlated motion of an electron-hole pair coupled by an effective electron-electron interaction. First applications to real systems (Cooke 1973; Cooke 1976) employed a tight-binding description. Reasonable agreement with experiment throughout the Brillouin zone was obtained for the spin-wave dispersion of $3d$ ferromagnets. With a similar approach, Tang et al. (1998) examined the spin dynamics in ultrathin ferromagnetic films on nonmagnetic substrates.

Around the turn of the millennium, the first calculations based on ab initio electronic structure methods were carried out: Karlsson and Aryasetiawan (2000) employed MBPT but used a local model potential with an adjustable parameter instead of the nonlocal electron-hole interaction. Savrasov (1998), Buczek et al. (2009), Lounis et al. (2010), and Rousseau et al. (2012) performed calculations

within time-dependent density functional theory (TDDFT). Kotani and van Schilf-
gaarde (2008) studied (anti-)ferromagnets based on quasiparticle self-consistent *GW*
calculations (Faleev et al. 2004), where the effective interaction was determined
from a magnetic sum rule. In 2010, Şaşıoğlu et al. (2010) reported a treatment
within MBPT where the screened Coulomb interaction was explicitly calculated
from the random-phase approximation (RPA) rather than using a model potential or
a parameter fixed by a sum rule. Still an ad hoc scaling of the screened interaction
was required to fulfill the Goldstone condition. Good agreement of the calculated
magnon dispersions in $3d$ ferromagnets with experiment was achieved in both
approaches, MBPT and TDDFT.

In the present chapter, we give a detailed presentation of the theoretical many-
body treatment of spin excitations within MBPT. We adopt a very general viewpoint
that encompasses excitations with and without a spin transfer, so neutral excita-
tions – e.g., excitons, which play an important role in optical absorption – will
appear as a special case. In Şaşıoğlu et al. (2010), a practical computational scheme
was developed to study excitation spectra of magnetic materials from first principles,
in close relation to the formalism of Aryasetiawan and Karlsson (1999). To study
collective magnon excitations, we include vertex corrections in the form of ladder
diagrams, which describe the coupling of electrons and holes with opposite spins via
the screened Coulomb interaction. In analogy to the many-body $T$ matrix defined by
Strinati (1988) for optical absorption, we use the same term for the corresponding
quantity that appears in the Green-function formalism for the transverse magnetic
response function. In order to reduce the numerical cost for the calculation of the
four-point $T$ matrix, we exploit a transformation to maximally localized Wannier
functions (MLWFs), which provide a more efficient basis to study local correlations
than extended Bloch states (Marzari and Vanderbilt 1997; Souza et al. 2001;
Freimuth et al. 2008). Our implementation is based on the full-potential linearized
augmented plane-wave (FLAPW) method.

Section 2 gives a detailed account of the theoretical framework. The numerical
implementation is described in Sect. 3. In particular, we discuss how the magnetic
response function can be calculated for any Bloch vector, even if this Bloch vector
is not an element of the **k**-point set. This allows the calculation of smooth dispersion
curves while keeping the **k**-point set small. Section 4 is devoted to the discussion of
the violation of the Goldstone theorem. This theorem stipulates the existence of an
acoustic magnon branch with vanishing excitation energy in the long-wavelength
limit. In numerical calculations, the excitation energy often remains finite in this
limit. We show – both numerically and mathematically – that this violation is due
to an inconsistency in the choice of the single-particle Green function. In Sect. 5,
we present illustrative magnetic excitation spectra obtained for the elementary
ferromagnets bcc Fe, fcc Co, and fcc Ni. For more practical applications of this
method, we refer the reader to Friedrich et al. (2014) and Şaşıoğlu et al. (2010,
2013). In Sect. 6, we summarize our conclusions. Unless otherwise indicated,
Hartree atomic units are used throughout.

## 2    Theory

When a many-electron system is perturbed by a time-dependent external **B** field, originating, for example, from a neutron beam impinging on a magnetic sample, the system reacts by a change of the electronic density. This electronic redistribution is different for spin-up and spin-down electrons, since the **B** field couples to the electrons' spin, and, to a first approximation, it is the spin density that changes, while the total density remains the same. Therefore, we consider the magnetic response function

$$R^{ij}(\mathbf{r}_1 t_1, \mathbf{r}_2 t_2) = \frac{\delta \sigma^i(\mathbf{r}_1 t_1)}{\delta B^j(\mathbf{r}_2 t_2)} \,, \tag{1}$$

which gives the linear change of the spin density at the position $\mathbf{r}_1$ and time $t_1$ with respect to changes in the external **B** field at $\mathbf{r}_2$ and $t_2$. Here, $\sigma^i$ and $B^j$ are the vector components ($i, j = x, y, z$) of the spin density $\boldsymbol{\sigma}$ and the **B** field. We additionally allow for $i = 0$ and $j = 0$, where $\sigma^0$ is the total electronic density and $B^0$ is an external scalar potential. For example, $R^{00} = \delta\sigma^0/\delta B^0$ then corresponds to the density response function. Equation (1) thus defines a $4 \times 4$ tensor of response functions that are all nonzero in general.

The magnetic response function exhibits resonances (analytical poles) at the spin excitation energies of the unperturbed many-body system, corresponding to the eigen oscillations of the spin system. These "eigen oscillations" are, in general, not discrete and show a spectral distribution given by the imaginary part of the magnetic response function. In order to capture all possible oscillations of the spin system, Eq. (1) defines a microscopic response function in the sense that the perturbing field – and also the response of the electronic (spin) density – can have an arbitrary shape in space. In particular, it can exhibit any wavelength down to interatomic distances. Its determination requires an ab initio description of the electronic structure and a high-level quantum mechanical treatment of the correlated motion of the electrons.

We employ a method based on MBPT similar to the one of Aryasetiawan and Karlsson (1999). However, we do not employ Matsubara frequencies but a formulation at absolute zero that yields the magnetic excitation spectra directly for real frequencies. An implementation within an all-electron Wannier-function formulation was published in Şaşıoğlu et al. (2010) and Friedrich et al. (2014). In the following, we develop the theory in detail.

To simplify the notation, space and time arguments $\mathbf{r}_1 t_1, \mathbf{r}_2 t_2, \ldots$ are abbreviated by the corresponding index $1, 2, \ldots$. The ground-state spin density distribution is given by the expectation value of the spin density operator

$$\sigma^i(1) = \sum_{\alpha,\beta} \sigma^i_{\alpha\beta} \langle \Psi_0 | \psi^\dagger_\alpha(1) \psi_\beta(1) | \Psi_0 \rangle \,, \tag{2}$$

where $\psi_\alpha(1)$ $[\psi_\alpha^\dagger(1)]$ is an annihilation (creation) field operator of an electron with spin $\alpha$ and $\sigma_{\alpha\beta}^i$ are the elements of the Pauli spin matrices for $i = x, y, z$. The expectation value is taken with respect to the interacting many-body ground state $|\Psi_0\rangle$. For example, for the spin density in $z$ direction, the operator is simply $\psi_\uparrow^\dagger(1)\psi_\uparrow(1) - \psi_\downarrow^\dagger(1)\psi_\downarrow(1)$, and the spin density is given by the difference of the spin-up and spin-down number densities. When considering transversal spin-wave excitations later on, the spin Pauli matrices for $i = x$ and $i = y$ will become relevant, but until then the derivation is general. The case $i = 0$ can be taken into account by defining $\sigma_{\alpha\beta}^0 = \delta_{\alpha\beta}$. We rewrite Eq. (2) as

$$\sigma^i(1) = -i \sum_{\alpha,\beta} \sigma_{\beta\alpha}^i G_{\alpha\beta}(11^+) \qquad (3)$$

where $1^+ = \mathbf{r}_1 t_1 + \eta$ with a positive infinitesimal time $\eta$ and $G_{\alpha\beta}(12) = -i\langle\Psi_0|\mathcal{T}[\psi_\alpha(1)\psi_\beta^\dagger(2)]|\Psi_0\rangle$ is the interacting single-particle Green function with the time-ordering operator $\mathcal{T}$ (Mahan 2000), which orders the field operators chronologically from right to left. (A factor $-1$ has to be multiplied for each permutation of field operators.)

We are now in the position to prove that $R^{ij}$ can be written as a spin-spin correlation function. First the Green function is expressed in the interaction picture

$$G_{\alpha\beta}(12) = -i\langle\Psi_0^I(\infty)|U^I(\infty, 1)\psi_\alpha^I(1)U^I(1, 2)\psi_\beta^{I\dagger}(2)U^I(2, -\infty)|\Psi_0\rangle$$

$$= -i\frac{\langle\Psi_0|U^I(\infty, 1)\psi_\alpha^I(1)U^I(1, 2)\psi_\beta^{I\dagger}(2)U^I(2, -\infty)|\Psi_0\rangle}{\langle\Psi_0|U(\infty, -\infty)|\Psi_0\rangle}, \qquad (4)$$

where we have assumed that $t_1 > t_2$, that the Heisenberg state $|\Psi_0\rangle$ is identical to $|\Psi_0^I(-\infty)\rangle$, and that $|\Psi_0^I(\infty)\rangle$ differs from $|\Psi_0\rangle$ only by a phase factor. The time evolution operator depends only on the time arguments and fulfills the Tomonaga-Schwinger equation

$$i\frac{\partial}{\partial t}U^I(t, t') = H^I(t)U^I(t, t') \qquad (5)$$

with the Zeeman term $H^I(t') = \sum_{j,\alpha,\beta} \sigma_{\alpha\beta}^j \int B^j(1)\psi_\alpha^{I\dagger}(1)\psi_\beta^I(1)d1$. Using the solution $U^I(t, t') = \mathcal{T}\exp\left[-i\int_{t'}^t H_j^I(t'')dt''\right]$, one can show that

$$\frac{\delta U^I(t, t')}{\delta B^j(3)} = \begin{cases} -iU^I(t, t_3)\hat{\sigma}^j(3)U^I(t_3, t') & \text{if } t < t_3 < t' \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

with $\hat{\sigma}^j(3) = \sum_{\alpha,\beta} \sigma_{\alpha\beta}^j \psi_\alpha^{I\dagger}(3)\psi_\beta^I(3)$. This expression replaces the corresponding $U^I$ when Eq. (4) is differentiated. Then, transforming back to the Heisenberg picture,

including the case $t_1 < t_2$, and inserting the result into Eqs. (1) and (3) yields the spin-spin correlation function

$$R^{ij}(12) = -i\langle\Psi_0|\mathcal{T}[\hat{\sigma}^{i\prime}(1)\hat{\sigma}^{j\prime}(2)]|\Psi_0\rangle \tag{7}$$

with $\hat{\sigma}^{i\prime}(1) = \hat{\sigma}^i(1) - \sigma^i(1)$.

The reformulation with the Green function makes Eq. (3) amenable to a treatment within MBPT. The equation of motion of the Green function, the Dyson equation, in the presence of a **B** field reads

$$\left[i\frac{\partial}{\partial t_1} + \frac{1}{2}\nabla_{\mathbf{r}_1}^2 - V^{\text{ext}}(\mathbf{r}_1)\right]G_{\alpha\beta}(12) - \left[\sum_i \sigma_{\alpha\beta}^i B^i(1)\right]G_{\alpha\beta}(12)$$

$$- \sum_\gamma \int M_{\alpha\gamma}(13)G_{\gamma\beta}(32)d3 = \delta(12)\delta_{\alpha\beta}\,, \tag{8}$$

from which we can directly identify the inverse of the Green function

$$G_{\alpha\beta}^{-1}(12) = \left[i\frac{\partial}{\partial t_1} + \frac{1}{2}\nabla_{\mathbf{r}_1}^2 - V^{\text{ext}}(\mathbf{r}_1)\right]\delta(12)\delta_{\alpha\beta} - \left[\sum_i \sigma_{\alpha\beta}^i B^i(1)\right]\delta(12) - M_{\alpha\beta}(12) \tag{9}$$

with the external potential $V^{\text{ext}}(\mathbf{r})$ and the delta function $\delta(12) = \delta(\mathbf{r}_1 - \mathbf{r}_2)\delta(t_1 - t_2)$. We have assumed the **B** field to incorporate a factor $g_e\mu_B/2$ ($g_e/4$ in atomic units) with the electron spin $g$-factor $g_e$ and the Bohr magneton $\mu_B$ so that $\mathbf{B}\cdot\hat{\sigma}$ is the Zeeman term of the Hamiltonian.

In solids, the orbital magnetic moment is usually strongly quenched, which is why we neglect the coupling of the **B** field to the orbital motion. The mass operator

$$M_{\alpha\beta}(12) = V^{\text{H}}(1)\delta(12)\delta_{\alpha\beta} + \Sigma_{\alpha\beta}(12) \tag{10}$$

accounts for the electron-electron interaction. It embodies the Hartree potential

$$V^{\text{H}}(1) = \int n(2)v(21)d2 = -i\sum_\alpha \int G_{\alpha\alpha}(22^+)v(21)d2 \tag{11}$$

with the bare Coulomb interaction $v(12) = \delta(t_1 - t_2)/|\mathbf{r}_1 - \mathbf{r}_2|$ and the self-energy $\Sigma_{\alpha\beta}(12)$, a time-dependent nonlocal potential that incorporates all many-body exchange and correlation effects of the electronic system. The self-energy is the most complex quantity in Eq. (9), and its exact form is unknown. We employ the $GW$ approximation (Hedin 1965)

$$\Sigma_{\alpha\beta}(12) = iG_{\alpha\beta}(12)W(1^+2)\,, \tag{12}$$

where

$$W(12) = v(12) + \iint v(13)P(34)W(42)\,d3\,d4 \qquad (13)$$

is the screened interaction, which is the effective potential at 1 created by a unit charge at 2 (first term) and the induced charge cloud forming around the unit charge (second term). The screened interaction is a dynamical (i.e., time-dependent) quantity, since the screening process requires the electrons to move, and this process takes time. Equation (13) is formally exact. The approximation consists in the choice of the polarization function, for which we use the RPA

$$P(12) = -\mathrm{i} \sum_{\alpha,\beta} G_{\alpha\beta}(12)G_{\beta\alpha}(21^+)\,. \qquad (14)$$

We note that Eq. (12) is defined with the self-consistently renormalized Green function, formally corresponding to a fully self-consistent solution of Hedin's equations (Hedin 1965) where the vertex function is approximated by $\Gamma(12;3) = \delta(12)\delta(13)$. This will become important in Sect. 4.

The derivative of $G$ can be related to that of $G^{-1}$ by differentiating both sides of $\sum_\gamma G_{\alpha\gamma}(13)G_{\gamma\beta}^{-1}(32)d3 = \delta(12)\delta_{\alpha\beta}$, yielding

$$\frac{\delta G_{\alpha\beta}(12)}{\delta B^j(3)} = -\sum_{\gamma,\delta} \iint G_{\alpha\gamma}(14)\frac{\delta G_{\gamma\delta}^{-1}(45)}{\delta B^j(3)}G_{\delta\beta}(52)d4\,d5\,. \qquad (15)$$

Through the derivative of Eq. (9)

$$\frac{\delta G_{\alpha\beta}^{-1}(12)}{\delta B^j(3)} = -\sigma_{\alpha\beta}^j\delta(13)\delta(12) - \frac{\delta M_{\alpha\beta}(12)}{\delta B^j(3)} \qquad (16)$$

and Eqs. (10), (11), and (12), the right-hand side of Eq. (15) can be expressed in terms of $\delta G/\delta B$, and successive insertion will lead to an infinite series expansion. Before we do this step, we have to find a suitable expression for the second term of $\delta\Sigma/\delta B = \mathrm{i}(\delta G/\delta B)W + \mathrm{i}G(\delta W/\delta B)$. Differentiating Eq. (13), solving for the derivative of $W$, and using Eq. (14) gives

$$\frac{\delta W(12)}{\delta B^j(3)} = -\mathrm{i} \sum_{\alpha,\beta} \iint W(14)$$

$$\left[\frac{\delta G_{\alpha\beta}(45)}{\delta B^j(3)}G_{\beta\alpha}(54^+) + G_{\alpha\beta}(45)\frac{\delta G_{\beta\alpha}(54^+)}{\delta B^j(3)}\right] W(52)\,d4\,d5\,. \qquad (17)$$
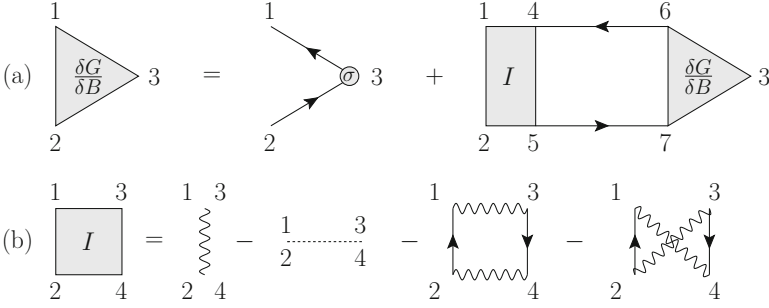
**Fig. 1** Diagrammatic representations of (**a**) Eq. (18) and (**b**) Eq. (19). Successive reinsertion in (**a**) produces an infinite series of diagrams for $\delta G/\delta B$ and, hence, for $R$. Arrows represent the renormalized Green function, the dotted line the bare, and the wiggly line the screened interaction. (We use the convention that each interaction line carries a factor i.) The spin-flip operator, mathematically described by a Pauli matrix, is shown as $\sigma$

Now, we can combine everything and obtain

$$\frac{\delta G_{\alpha\beta}(12)}{\delta B^j(3)} = \sum_{\gamma,\delta}\left[\sigma^j_{\gamma\delta}G_{\alpha\gamma}(13)G_{\delta\beta}(32^+)\right. \tag{18}$$

$$\left. + \sum_{\varepsilon,\zeta}\iiiint G_{\alpha\gamma}(14)G_{\delta\beta}(52^+)I_{\gamma\delta,\varepsilon\zeta}(45,67)\frac{\delta G_{\varepsilon\zeta}(67)}{\delta B^j(3)}d4\,d5\,d6\,d7\right]$$

with the (irreducible) interaction kernel

$$I_{\alpha\beta,\gamma\delta}(12,34) = \mathrm{i}\left[W(1^+2)\delta(13)\delta(24)\delta_{\alpha\gamma}\delta_{\beta\delta} - v(13)\delta_{\alpha\beta}\delta_{\gamma\delta}\delta(12)\delta(3^+4)\right]$$
$$+ G_{\alpha\beta}(12)\left[W(1^+3)G_{\delta\gamma}(43^+)W(42)+W(1^+4)G_{\delta\gamma}(43^+)W(32)\right]. \tag{19}$$

Equation (18) defines an infinite series expansion for $\delta G/\delta B$ as shown diagrammatically in Fig. 1a. By virtue of the Eqs. (1) and (3), a corresponding series expansion is obtained for $R^{ij}(12)$.

For $i = j = 0$, the above formulas lead to the Bethe-Salpeter equation for optical absorption. The case $i = 0 \neq j$ (or vice versa) describes the coupling of electronic spin and charge. The equations simplify in the absence of spin-orbit coupling. The Green function is then diagonal in spin space, and the inner spin summations in Eq. (18) disappear. We now discuss some specific cases:

**Nonmagnetic case:** We have $G_{\alpha\beta}(12) = \delta_{\alpha\beta}G(12)$, i.e., spin-up and spin-down Green functions are identical. As a result, all degrees of freedom (0, $x$, $y$, and $z$) decouple and $R^{00} = R^{zz}$ and $R^{xx} = R^{yy}$.

**Collinear (anti-)ferromagnetism:** The Green function is still diagonal in spin space, but the two components are different, i.e., $G_{\alpha\beta}(12) = \delta_{\alpha\beta}G_\alpha(12)$. Then, the degrees of freedom 0 and $z$ as well as $x$ and $y$ couple. In particular, we get a coupling of the spin and charge degrees of freedom in $R^{0z}$ (and $R^{z0}$). The diagonal elements $R^{00}$ and $R^{zz}$ are the density response function and the longitudinal spin susceptibility, respectively. They are decoupled from the functions $R^{ij}$ with $i, j \in \{x, y\}$, which give rise to the transverse spin susceptibility. In the latter case, the interaction kernel reduces to the first term $I = iW$.

**Spin-orbit coupling:** In the presence of spin-orbit coupling, the Green function acquires spin-off-diagonal elements. Then all degrees of freedom couple in general and all terms in Eq. (19) and all response functions defined in Eq. (1) must be taken into account.

In the following, we restrict ourselves to the case of spin excitations in a ferromagnet with a collinear magnetic ground state without spin-orbit coupling. For a spin polarization along the $z$ axis, these are generated by an oscillating magnetic field in the $xy$ plane. In particular, in inelastic neutron scattering experiments, the incoming neutron beam is circularly polarized, creating a magnetic field whose $y$ component exhibits a phase shift of $\pi/2$ with respect to the $x$ component (Lowde et al. 1983).

It is instructive to discuss the coupling of the electron's spin to the $\mathbf{B}$ field in terms of semiclassical physics. Without the perturbing field, the electron spin $\mathbf{S}$ precesses around the Weiss (exchange) field $\mathbf{B}_0$ according to the equation of motion $\dot{\mathbf{S}} = \boldsymbol{\mu} \times \mathbf{B}_0$, where $\boldsymbol{\mu} = -g_e\mu_B\mathbf{S}/\hbar$ is the magnetic moment of the electron. (We write the formulas in SI units in this paragraph.) With the majority spin pointing in the positive $\mathbf{z}$ direction, the Weiss field points in the negative $\mathbf{z}$ direction, and $\omega_0 = g_e\mu_B\mathbf{B}_0/\hbar$ is the Larmor frequency of the precession. The two possible circular polarizations of the perturbing $\mathbf{B}$ field – right- and left-handed with respect to the $\mathbf{B}_0$ field – are given by $\mathbf{B}^\pm(t) = B^\pm \mathrm{Re}[(\hat{\mathbf{x}}\mp i\hat{\mathbf{y}})e^{-i\omega t}]$, respectively. The equation of motion in the presence of the perturbing field is $\dot{\mathbf{S}} = \boldsymbol{\mu} \times (\mathbf{B}_0 + \mathbf{B}^\pm)$. If $\mathbf{S}^\pm$ is defined as the spin vector seen in the coordinate system that rotates with the $\mathbf{B}^\pm$ field, we can write $\dot{\mathbf{S}} = \dot{\mathbf{S}}^\pm - \mathbf{S} \times \boldsymbol{\omega}$. Equating the right-hand sides of the last two equations and inserting the formulas for $\boldsymbol{\mu}$ and $\omega_0$ gives $\dot{\mathbf{S}}^\pm = \boldsymbol{\mu} \times \mathbf{B}^\pm + \mathbf{S} \times (\boldsymbol{\omega} - \boldsymbol{\omega}_0)$, which reduces to $\dot{\mathbf{S}}^\pm = \boldsymbol{\mu} \times \mathbf{B}^\pm$ if $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. This is the equation of motion of a spin precessing around $\mathbf{B}^\pm$, i.e., around a direction perpendicular to $\mathbf{B}_0$, making it possible to flip the spin of the electron. The frequency of this precession, the Rabi frequency, is $g_e\mu_B B/\hbar$. The condition that $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0 = g_e\mu_B\mathbf{B}_0/\hbar$ have the same orientation is fulfilled for the right-handed circular polarization, for which the relevant component is $B^+ = B_x + iB_y$ because $B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}} = B^+(\hat{\mathbf{x}} - i\hat{\mathbf{y}}) + B^-(\hat{\mathbf{x}} + i\hat{\mathbf{y}})$ with $B^- = B_x - iB_y$.

Therefore, we consider the transverse magnetic susceptibility

$$R^{+-}(12) = \frac{\delta\sigma^+(1)}{\delta B^+(2)} \tag{20}$$

in the following, where "$+$" and "$-$" refer to the Pauli matrices $\hat{\sigma}^+ = \hat{\sigma}^x + i\hat{\sigma}^y$ and $\hat{\sigma}^- = \hat{\sigma}^x - i\hat{\sigma}^y$ with the matrix representations

$$\sigma^+ = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \sigma^- = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \tag{21}$$

for the spin creation and annihilation operators. The Zeeman term of Eq. (9) can be written as

$$\sum_i \sigma_{\alpha\beta}^i B^i(1) = \frac{1}{2} \left[ \sigma_{\alpha\beta}^+ B^-(1) + \sigma_{\alpha\beta}^- B^+(1) \right] + \sigma_{\alpha\beta}^z B^z(1). \tag{22}$$

Because of the mixed products of the form $\sigma^+ B^-$ and $\sigma^- B^+$, the spin-spin correlation function Eq. (7) becomes

$$R^{+-}(12) = \frac{\delta\sigma^+(1)}{\delta B^+(2)} = -i\langle\Psi_0|\mathscr{T}[\hat{\sigma}^+(1)\hat{\sigma}^-(2)]|\Psi_0\rangle, \tag{23}$$

where we have used that $\langle\Psi_0|\hat{\sigma}^+(1)|\Psi_0\rangle = \langle\Psi_0|\hat{\sigma}^-(1)|\Psi_0\rangle = 0$ in a collinear magnetic system. This form of the spin-spin correlation function is intuitive: for a spin in the up channel to be probed, one has to flip its spin with the operator $\hat{\sigma}^-$ before flipping it back with $\hat{\sigma}^+$. Equation (23) also explains the notation $R^{+-}$.

With Eq. (22), a derivation that proceeds in analogy to above leads to

$$i\frac{\delta G_{\downarrow\uparrow}(12)}{\delta B^+(3)} = K_{\downarrow\uparrow}(12, 33) + \iint K_{\downarrow\uparrow}(12, 45)W(4^+5)i\frac{\delta G_{\downarrow\uparrow}(45)}{\delta B^j(3)}d4\,d5 \tag{24}$$

with the (uncorrelated) two-particle propagator

$$K_{\downarrow\uparrow}(12, 34) = iG_{\downarrow}(13)G_{\uparrow}(42^+) \tag{25}$$

($G_{\uparrow} = G_{\uparrow\uparrow}$, $G_{\downarrow} = G_{\downarrow\downarrow}$). We have used that (a) the mass operator Eq. (10) reduces to $\Sigma$ for $\alpha \neq \beta$ and (b) the second term in Eq. (19) vanishes, (c) as do the last two terms because the Green function is diagonal in spin space. Note that Eq. (24) can be obtained from Eq. (18) by setting $i = +$ and $j = -$ except for an additional factor $1/2$ from Eq. (22).

Equation (24) can be written as a matrix equation if we define a generalized four-point magnetic response function (formally giving the response of the spin density matrix with respect to changes of a nonlocal $\mathbf{B}$ field). To this end, we introduce

the auxiliary four-point magnetic response function defined by the Bethe-Salpeter equation

$$R_{\downarrow\uparrow}^{(4)}(12, 34) = K_{\downarrow\uparrow}(12, 34) + \iint K_{\downarrow\uparrow}(12, 56)W(5^+6)R_{\downarrow\uparrow}^{(4)}(56, 34)d5\,d6\,, \tag{26}$$

from which we deduce

$$R_{\downarrow\uparrow}^{(4)}(12, 33) = \mathrm{i}\frac{\delta G_{\downarrow\uparrow}(12)}{\delta B^+(3)} \tag{27}$$

$$R^{+-}(12) = -2R_{\downarrow\uparrow}^{(4)}(11, 22)\,. \tag{28}$$

The magnetic response function can, furthermore, be written as the sum of two terms

$$R^{+-}(12) = -2K_{\downarrow\uparrow}(11, 22) - 2\iiiint K_{\downarrow\uparrow}(11, 34)T_{\downarrow\uparrow}(34, 56)$$
$$K_{\downarrow\uparrow}(56, 22)d3\,d4\,d5\,d6\,, \tag{29}$$

where the $T$ matrix, which can be interpreted as a reducible interaction kernel, fulfills the equation

$$T_{\downarrow\uparrow}(12, 34) = W(1^+2)\delta(13)\delta(24) + \iint W(1^+2)K_{\downarrow\uparrow}(12, 56)T_{\downarrow\uparrow}(56, 34)d5\,d6\,. \tag{30}$$

If we approximate the renormalized Green function in Eq. (25) by the Kohn-Sham Green function, then the first term of Eq. (29) contains single-particle excitations between Kohn-Sham levels from one spin channel into the other. The second term describes the correlated motion of an electron-hole pair with opposite spins through the $T$ matrix Fig. 2; each ladder diagram stands for a series of scattering events. This correlated motion is responsible for the occurrence of collective spin excitations. The second term also renormalizes the Stoner excitations.

TDDFT (Runge and Gross 1984) is another method that allows one to calculate the magnetic response function from first principles. In this theory, Eq. (3) is written in terms of the Kohn-Sham Green function $G_{\alpha\beta}^{\mathrm{KS}}(12)$ instead of $G_{\alpha\beta}(12)$. This
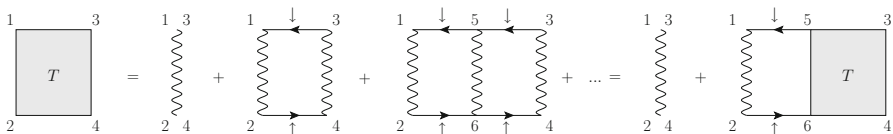


**Fig. 2** Diagrammatic representation of the $T_{\downarrow\uparrow}(12, 34)$ matrix consisting of an infinite series of ladder diagrams, each wiggly line (screened interaction $W$) representing a "rung" of the ladder. The indices 5 and 6 denote integration variables

is not an approximation since the Kohn-Sham system (Kohn and Sham 1965) is constructed to yield the *exact* electronic spin density of the interacting system (provided that the exact exchange-correlation functional is employed). A derivation, similar to the one above with $\Sigma$ in Eq. (10) replaced by the exchange-correlation potential $v_{\alpha\beta}^{\text{xc}}(1)$, then yields an expression for the magnetic response function with the interaction kernel

$$I_{\alpha\beta,\gamma\delta}^{\text{KS}}(12,34) = -\mathrm{i}\left[v(13)\delta_{\alpha\beta}\delta_{\gamma\delta} + f_{\alpha\beta,\gamma\delta}^{\text{xc}}(13)\right]\delta(12)\delta(3^+4)\,, \qquad (31)$$

where $f_{\alpha\beta,\gamma\delta}^{\text{xc}}(12) = \delta v_{\alpha\beta}^{\text{xc}}(1)/\delta\sigma_{\gamma\delta}(2)$ is the exchange-correlation kernel of TDDFT. This interaction kernel is effectively just a two-point function. Equation (26), hence, turns into a two-point matrix equation, which makes the implementation of the TDDFT approach much simpler and computationally less demanding. However, there are only few approximations available for $f_{\alpha\beta,\gamma\delta}^{\text{xc}}(12)$ to date, which, at that, cannot be systematically improved in contrast to the electronic self-energy. The most common approximation, the adiabatic local-density approximation, neglects time dependence and nonlocality altogether (and also requires the spin indices to be pairwise identical) so that $f_{\alpha\beta,\beta\alpha}^{\text{xc}}(1)$ becomes a local function. Still, several publications (Savrasov 1998; Buczek et al. 2009, 2010, 2011; Lounis et al. 2010, 2011) have demonstrated that the spin excitation spectra calculated within TDDFT are often in good agreement to results from MBPT and to experiment.

## 3   Implementation

In this section, we present the basics of a numerical implementation in the SPEX code (Friedrich et al. 2010). For more details, we refer the reader to Friedrich et al. (2014). The four-point quantities derived in the previous section are represented in a basis of Wannier functions. In Sect. 4, we will discuss and compare several mean-field systems as the reference noninteracting system. In the present section, we assume the Kohn-Sham solution (Kohn and Sham 1965) of density functional theory (DFT) (Hohenberg and Kohn 1964) be used. The corresponding Kohn-Sham equations are solved within the all-electron FLAPW method as implemented in the FLEUR code (http://www.flapw.de), which allows an accurate representation of the single-particle states $\varphi_{\mathbf{k}m}^{\alpha}(\mathbf{r})$, where $\mathbf{k}$ is the Bloch vector and $m$ the band index.

For a practical implementation, Eq. (26) is too complex because it contains quantities that, in general, depend on four points in space and on four time (or frequency) arguments (three if the Hamiltonian is time independent). A first simplification uses the fact that the spin-wave excitations are usually of low frequency, which motivates to replace the screened interaction by its static limit, i.e., $W(\mathbf{r}_1, \mathbf{r}_2; \omega) \rightarrow W(\mathbf{r}_1, \mathbf{r}_2; 0) = W(\mathbf{r}_1, \mathbf{r}_2)$, implying an instantaneous interaction in time, $W(\mathbf{r}_1 t_1, \mathbf{r}_2 t_2) = W(\mathbf{r}_1, \mathbf{r}_2)\delta(t_1 - t_2)$. For example, the two-particle propagator Eq. (25) then only depends on a single time (or frequency) argument because the delta function in the previous expression (and the contraction of

Eq. (28)) requires the argument pairs 12 and 34 to have the same time argument, i.e., $K_{\downarrow\uparrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_3, \mathbf{r}_4; \tau = t_1 - t_3)$ with $t_1 = t_2$ and $t_3 = t_4$. Equation (26) then simplifies to

$$R_{\downarrow\uparrow}^{(4)}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_3, \mathbf{r}_4; \omega) = K_{\downarrow\uparrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_3, \mathbf{r}_4; \omega) + \iint K_{\downarrow\uparrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_5, \mathbf{r}_6; \omega)$$

$$W(\mathbf{r}_5, \mathbf{r}_6) R_{\downarrow\uparrow}^{(4)}(\mathbf{r}_5, \mathbf{r}_6; \mathbf{r}_3, \mathbf{r}_4; \omega) \, d^3 r_5 \, d^3 r_6 \qquad (32)$$

in the frequency domain. We employ another approximation in that we allow electrons (and holes) to interact with each other only when they are located on the same atomic site, thus making use of the fact that the screened interaction is short range in metallic systems. This on-site approximation is not strictly necessary, but it greatly simplifies the implementation and enables fast calculations. Besides, it is a very good approximation for the systems studied here (Müller et al. 2016).

Wannier functions (Marzari and Vanderbilt 1997) are localized functions defined by linear combinations of the single-particle wave functions

$$w_{\mathbf{R}n}^{\alpha}(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} \sum_{m} U_{\mathbf{k}m,n}^{\alpha} \varphi_{\mathbf{k}m}^{\alpha}(\mathbf{r}) \,, \qquad (33)$$

where $n$ is an index counting the Wannier functions at the atomic site $\mathbf{R}$ and $N$ is the number of $\mathbf{k}$ points of an $N_x \times N_y \times N_z$ Monkhorst-Pack set (Monkhorst and Pack 1976) including $\mathbf{k} = \mathbf{0}$. The transformation matrix $U_{\mathbf{k}m,n}^{\alpha}$ is determined by minimizing the spread of the Wannier functions (Souza et al. 2001; Marzari and Vanderbilt 1997; Freimuth et al. 2008), under the condition that the Wannier functions are orthonormal with respect to integrations over the $N_x \times N_y \times N_z$ supercell (whereas the $\varphi_{\mathbf{k}m}^{\alpha}(\mathbf{r})$ are orthonormalized with respect to the unit cell). The sum over $m$ runs over a limited number of electronic bands (at least as many as the number of Wannier functions).

In the frequency domain, Eq. (25) becomes

$$K_{\downarrow\uparrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_3, \mathbf{r}_4; \omega) = \frac{i}{2\pi} \int_{-\infty}^{\infty} G_{\downarrow}(\mathbf{r}_1, \mathbf{r}_3; \omega') G_{\uparrow}(\mathbf{r}_4, \mathbf{r}_2; \omega' - \omega) d\omega'$$

$$= \frac{1}{N^2} \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \sum_{m}^{\text{occ.}} \sum_{m'}^{\text{unocc.}} \left\{ \frac{\varphi_{\mathbf{k}m}^{\downarrow}(\mathbf{r}_1) \varphi_{\mathbf{k}m}^{\downarrow*}(\mathbf{r}_3) \varphi_{\mathbf{k}'m'}^{\uparrow*}(\mathbf{r}_2) \varphi_{\mathbf{k}'m'}^{\uparrow}(\mathbf{r}_4)}{\omega + \varepsilon_{\mathbf{k}'m'}^{\uparrow} - \varepsilon_{\mathbf{k}m}^{\downarrow} - i\eta} \right.$$

$$\left. - \frac{\varphi_{\mathbf{k}'m'}^{\downarrow}(\mathbf{r}_1) \varphi_{\mathbf{k}'m'}^{\downarrow*}(\mathbf{r}_3) \varphi_{\mathbf{k}m}^{\uparrow*}(\mathbf{r}_2) \varphi_{\mathbf{k}m}^{\uparrow}(\mathbf{r}_4)}{\omega + \varepsilon_{\mathbf{k}m}^{\uparrow} - \varepsilon_{\mathbf{k}'m'}^{\downarrow} + i\eta} \right\} \qquad (34)$$

with a positive infinitesimal $\eta$, where we have used the Lehmann representation of the noninteracting Kohn-Sham Green function

$$G_\alpha(\mathbf{r}, \mathbf{r}'; \omega) = \frac{1}{N} \sum_\mathbf{k} \sum_m \frac{\varphi_{\mathbf{k}m}^\alpha(\mathbf{r})\varphi_{\mathbf{k}m}^{\alpha*}(\mathbf{r}')}{\omega - \varepsilon_{\mathbf{k}m}^\alpha + i\eta \operatorname{sgn}(\varepsilon_{\mathbf{k}m}^\alpha - \varepsilon_F)} \tag{35}$$

with the Fermi energy $\varepsilon_F$.

Each of the four vertices of $K_{\downarrow\uparrow}(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_3, \mathbf{r}_4; \omega)$ is now projected onto the Wannier basis defined by Eq. (33), which gives

$$
\begin{aligned}
K_{\mathbf{R}n_1\mathbf{R}n_2, \mathbf{R}'n_3\mathbf{R}'n_4}^{\downarrow\uparrow}(\omega) = \frac{1}{N^2} \sum_{\mathbf{k}m}^{\text{occ.}} \sum_{\mathbf{k}'m'}^{\text{unocc.}} \Bigg\{ & \frac{U_{\mathbf{k}m,n_1}^{\downarrow*} U_{\mathbf{k}m,n_3}^{\downarrow} U_{\mathbf{k}'m',n_2}^{\uparrow} U_{\mathbf{k}'m',n_4}^{\uparrow*}}{\omega + \varepsilon_{\mathbf{k}'m'}^{\uparrow} - \varepsilon_{\mathbf{k}m}^{\downarrow} - i\eta} e^{-i(\mathbf{k}'-\mathbf{k})(\mathbf{R}-\mathbf{R}')} \\
& - \frac{U_{\mathbf{k}'m',n_1}^{\downarrow*} U_{\mathbf{k}'m',n_3}^{\downarrow} U_{\mathbf{k}m,n_2}^{\uparrow} U_{\mathbf{k}m,n_4}^{\uparrow*}}{\omega + \varepsilon_{\mathbf{k}m}^{\uparrow} - \varepsilon_{\mathbf{k}'m'}^{\downarrow} + i\eta} e^{i(\mathbf{k}'-\mathbf{k})(\mathbf{R}-\mathbf{R}')} \Bigg\},
\end{aligned}
\tag{36}
$$

where the on-site approximation has been used to set $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{R}$ and $\mathbf{R}_3 = \mathbf{R}_4 = \mathbf{R}'$. This expression only depends on $\Delta\mathbf{R} = \mathbf{R} - \mathbf{R}'$, and a lattice Fourier transformation yields

$$
\begin{aligned}
K_{n_1n_2, n_3n_4}^{\downarrow\uparrow}(\mathbf{q}, \omega) &= \sum_{\Delta\mathbf{R}} K_{\mathbf{R}n_1\mathbf{R}n_2, \mathbf{R}-\Delta\mathbf{R}n_3\mathbf{R}-\Delta\mathbf{R}n_4}^{\downarrow\uparrow}(\omega) e^{-i\mathbf{q}\cdot\Delta\mathbf{R}} \\
&= \frac{1}{N} \sum_\mathbf{k} \sum_m^{\text{occ.}} \sum_{m'}^{\text{unocc.}} \Bigg\{ \frac{U_{\mathbf{q}+\mathbf{k}m,n_1}^{\downarrow*} U_{\mathbf{q}+\mathbf{k}m,n_3}^{\downarrow} U_{\mathbf{k}m',n_2}^{\uparrow} U_{\mathbf{k}m',n_4}^{\uparrow*}}{\omega + \varepsilon_{\mathbf{k}m'}^{\uparrow} - \varepsilon_{\mathbf{q}+\mathbf{k}m}^{\downarrow} - i\eta} \\
&\qquad - \frac{U_{\mathbf{q}+\mathbf{k}m',n_1}^{\downarrow*} U_{\mathbf{q}+\mathbf{k}m',n_3}^{\downarrow} U_{\mathbf{k}m,n_2}^{\uparrow} U_{\mathbf{k}m,n_4}^{\uparrow*}}{\omega + \varepsilon_{\mathbf{k}m}^{\uparrow} - \varepsilon_{\mathbf{q}+\mathbf{k}m'}^{\downarrow} + i\eta} \Bigg\}.
\end{aligned}
\tag{37}
$$

We use the tetrahedron method (Rath and Freeman 1975) for the $\mathbf{k}$ summation.

From this equation, it is clear that if $\mathbf{q}$ and $\mathbf{k}$ are elements of the $\mathbf{k}$-point set, then $\mathbf{q} + \mathbf{k}$ must be an element of the set, too. The Monkhorst-Pack grid fulfills this condition. On the other hand, this condition limits the number of $\mathbf{q}$ points at which a spin excitation spectrum can be calculated to the relatively few points of the $\mathbf{k}$-point set. In order to evaluate $K$ (and $R$) at an arbitrary Bloch vector $\mathbf{q}$, which would enable the calculation of smooth spin-wave dispersion curves, we have to introduce an auxiliary set of Wannier functions with a suitable Bloch character

$$\tilde{w}_{\mathbf{R}n}^\alpha(\mathbf{r}) = \frac{1}{N} \sum_\mathbf{k} e^{-i(\mathbf{k}+\mathbf{q})\cdot\mathbf{R}} \sum_m U_{\mathbf{k}+\mathbf{q}m,n}^\alpha \varphi_{\mathbf{k}+\mathbf{q}m}^\alpha(\mathbf{r}), \tag{38}$$

where the transformation matrices are distinguished from the ones used in Eq. (33) by the Bloch vectors $\mathbf{k} + \mathbf{q}$, which are generally not elements of the original $\mathbf{k}$-point set. With this definition, Eq. (37) remains valid, but it has to be taken into account

that the transformation matrices $U^\alpha_{\mathbf{q}+\mathbf{k}m,n}$ and $U^\alpha_{\mathbf{k}m,n}$ now belong to two different sets of Wannier functions.

The second quantity that we need for solving Eq. (26) is the screened interaction. We define its Wannier representation by

$$W^{\downarrow\uparrow}_{\mathbf{R}n_1\mathbf{R}n_2,\mathbf{R}n_3\mathbf{R}n_4}(\omega) = \frac{N}{\mathcal{N}}\iint w^{\downarrow*}_{\mathbf{R}n_1}(\mathbf{r})w^{\downarrow}_{\mathbf{R}n_3}(\mathbf{r})W(\mathbf{r},\mathbf{r}';\omega)w^{\uparrow}_{\mathbf{R}n_2}(\mathbf{r}')w^{\uparrow*}_{\mathbf{R}n_4}(\mathbf{r}')d^3r\,d^3r'\,,$$
(39)

where $\mathcal{N}(\mathcal{N}/N)$ is the infinite number of unit cells (supercells). The prefactor is required to avoid double counting because the integrations extend over the whole infinite space. Inserting Eq. (33) gives

$$W^{\downarrow\uparrow}_{\mathbf{R}n_1\mathbf{R}n_2,\mathbf{R}n_3\mathbf{R}n_4}(\omega) = \frac{1}{N^3}\sum_{\mathbf{k},\mathbf{k}',m_1,m_2,m_3,m_4}U^{\downarrow*}_{\mathbf{k}+\mathbf{k}''m_1,n_1}U^{\downarrow}_{\mathbf{k}m_3,n_3}U^{\uparrow}_{\mathbf{k}'+\mathbf{k}''m_2,n_2}U^{\uparrow*}_{\mathbf{k}'m_4,n_4}$$

$$\times\frac{1}{\mathcal{N}}\iint\varphi^{\downarrow*}_{\mathbf{k}+\mathbf{k}''m_1}(\mathbf{r})\varphi^{\downarrow}_{\mathbf{k}m_3}(\mathbf{r})W(\mathbf{r},\mathbf{r}';\omega)\varphi^{\uparrow}_{\mathbf{k}'+\mathbf{k}''m_2}(\mathbf{r}')$$

$$\varphi^{\uparrow*}_{\mathbf{k}'m_4}(\mathbf{r}')d^3r\,d^3r'\,,$$
(40)

where it has been used that $W(\mathbf{r},\mathbf{r}';\omega)$ is diagonal in $\mathbf{k}$. The evaluation of the double integral, which, together with the prefactor $1/\mathcal{N}$, is finite, is discussed elsewhere (Friedrich et al. 2009). Since Eq. (40) is independent of $\mathbf{R}$, we may write $W^{\downarrow\uparrow}_{n_1n_2,n_3n_4}(\omega)$.

We are now in the position to formulate the Bethe-Salpeter equation (Eq. (32)) in the Wannier basis

$$R^{(4)\downarrow\uparrow}_{\mathbf{R}n_1\mathbf{R}n_2,\mathbf{R}'n_3\mathbf{R}'n_4}(\omega) = K^{\downarrow\uparrow}_{\mathbf{R}n_1\mathbf{R}n_2,\mathbf{R}'n_3\mathbf{R}'n_4}(\omega) + \sum_{\mathbf{R}''}\sum_{n_5,n_6,n_7,n_8}K^{\downarrow\uparrow}_{\mathbf{R}n_1\mathbf{R}n_2,\mathbf{R}''n_5\mathbf{R}''n_6}(\omega)$$

$$\times W^{\downarrow\uparrow}_{n_5n_6,n_7n_8}(0)R^{(4)\downarrow\uparrow}_{\mathbf{R}''n_7\mathbf{R}''n_8,\mathbf{R}'n_3\mathbf{R}'n_4}(\omega)\,.$$
(41)

As $K$ (and hence also $R$) depends only on the difference $\Delta\mathbf{R} = \mathbf{R} - \mathbf{R}'$, we can insert the lattice Fourier transformations

$$K^{\downarrow\uparrow}_{\mathbf{R}n_1\mathbf{R}n_2,\mathbf{R}'n_3\mathbf{R}'n_4}(\omega) = \frac{1}{N}\sum_{\mathbf{q}}e^{i\mathbf{q}\cdot(\mathbf{R}-\mathbf{R}')}K^{\downarrow\uparrow}_{n_1n_2,n_3n_4}(\mathbf{q},\omega)\,,$$
(42)

analogously for $R^{(4)}$, and obtain

$$R^{(4)\downarrow\uparrow}_{n_1n_2,n_3n_4}(\mathbf{q},\omega) = K^{\downarrow\uparrow}_{n_1n_2,n_3n_4}(\mathbf{q},\omega) + \sum_{n_5,n_6,n_7,n_8}K^{\downarrow\uparrow}_{n_1n_2,n_5n_6}(\mathbf{q},\omega)$$

$$\times W^{\downarrow\uparrow}_{n_5n_6,n_7n_8}(0)R^{(4)\downarrow\uparrow}_{n_7n_8,n_3n_4}(\mathbf{q},\omega)\,.$$
(43)

This is a matrix equation in pairs of Wannier indices. Since Wannier functions are orthonormal, their pairs are orthonormal, too. The summations on the right-hand side are mere matrix multiplications, and we can formally solve Eq. (43) for $R^{(4)}$

$$R^{(4)\downarrow\uparrow}(\mathbf{q}, \omega) = [\mathbb{1} - K^{\downarrow\uparrow}(\mathbf{q}, \omega)W^{\downarrow\uparrow}(0)]^{-1}K^{\downarrow\uparrow}(\mathbf{q}, \omega). \tag{44}$$

Using Eq. (28), we obtain the physically relevant magnetic response function from the matrix elements of $R^{(4)\downarrow\uparrow}(\mathbf{q}, \omega)$ by

$$R^{+-}(\mathbf{r}, \mathbf{r}'; \omega) = -\frac{2}{N} \sum_{\mathbf{q}} \sum_{n_1, n_2, n_3, n_4} R^{(4)\downarrow\uparrow}_{n_1 n_2, n_3 n_4}(\mathbf{q}, \omega)\Omega^{\downarrow\uparrow}_{\mathbf{q}n_1 n_2}(\mathbf{r})\Omega^{\downarrow\uparrow*}_{\mathbf{q}n_3 n_4}(\mathbf{r}'). \tag{45}$$

where $\Omega^{\downarrow\uparrow}_{\mathbf{q}n_1 n_2}(\mathbf{r})$ are the lattice Fourier transforms of the Wannier products $\Omega^{\alpha\beta}_{\mathbf{R}, n_1 n_2}(\mathbf{r}) = w^{\alpha}_{\mathbf{R}n_1}(\mathbf{r})w^{\beta*}_{\mathbf{R}n_2}(\mathbf{r})$, i.e.,

$$\Omega^{\alpha\beta}_{\mathbf{q}n_1 n_2}(\mathbf{r}) = \sum_{\mathbf{R}} \Omega^{\alpha\beta}_{\mathbf{R}n_1 n_2}(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{R}} = \frac{1}{N} \sum_{\mathbf{k}} w^{\alpha}_{\mathbf{q}+\mathbf{k}n_1}(\mathbf{r})w^{\beta*}_{\mathbf{k}n_2}(\mathbf{r}) \tag{46}$$

with the Wannier Bloch functions $w^{\alpha}_{\mathbf{k}n}(\mathbf{r}) = \sum_m U^{\alpha}_{\mathbf{k}m, n}\varphi^{\alpha}_{\mathbf{k}m}(\mathbf{r})$. If we use the matrix representation of $K^{\downarrow\uparrow}(\mathbf{q}, \omega)$ in Eq. (45), we obtain the bare susceptibility, i.e., the fictitious magnetic response function of the noninteracting reference system:

$$K^{+-}(\mathbf{r}, \mathbf{r}'; \omega) = -2K^{\downarrow\uparrow}(\mathbf{r}, \mathbf{r}; \mathbf{r}', \mathbf{r}'; \omega) \tag{47}$$

$$= -\frac{2}{N} \sum_{\mathbf{q}} \sum_{n_1, n_2, n_3, n_4} K^{\downarrow\uparrow}_{n_1 n_2, n_3 n_4}(\mathbf{q}, \omega)\Omega^{\downarrow\uparrow}_{\mathbf{q}n_1 n_2}(\mathbf{r})\Omega^{\downarrow\uparrow*}_{\mathbf{q}n_3 n_4}(\mathbf{r}'). \tag{48}$$

Of course, $R^{+-}(\mathbf{r}, \mathbf{r}'; \omega)$ is still a very complex quantity: it is nonlocal in space, it shows a frequency dependence, and it has a real and an imaginary part. The spectrum measured in neutron scattering experiments, for example, can be extracted from $R^{+-}(\mathbf{r}, \mathbf{r}'; \omega)$ by projecting its imaginary part from left and right to a plane wave $e^{i\mathbf{q}\cdot\mathbf{r}}$ giving Im $R^{+-}(\mathbf{q}, \omega)$ (Lowde et al. 1983). Sharp peaks in this function correspond to collective spin excitations, the spin waves, with wavevector $\mathbf{q}$ and frequency $\omega$. Plotting the respective $\omega$ values against $\mathbf{q}$ yields the dispersion relation of the spin-wave mode. We note that another possibility is to perform a normal mode analysis of the imaginary part of $R^{+-}$, in matrix notation Im $R^{+-} = (R^{+-} - R^{+-\dagger})/(2i)$.

## 4    Goldstone Condition

The Goldstone theorem states that the spontaneously broken spin-rotation symmetry in ferromagnetic materials leads to the appearance of a gapless magnon dispersion curve, i.e., the excitation energy vanishes in the limit $\mathbf{q} \to \mathbf{0}$. This has a very simple

physical explanation. The generating field is a magnetic field perpendicular to the ferromagnetic spin alignment, and it is commensurate with the unit cell in the limit $\mathbf{q} \rightarrow \mathbf{0}$. If the $\mathbf{B}$ field has a suitable shape in real space, it can act to rotate all electron spins collectively toward the field direction, i.e., the electron spins point in the same direction at all times. In the absence of spin-orbit coupling, this rigid rotation takes place without a cost of energy; hence $\omega = 0$. The pole at $\omega = 0$ corresponds to the long-wavelength limit of the acoustic magnon branch (Moriya 1985), identified as the *Goldstone mode*.

A rigid rotation of all spins does not change any relative spin alignment. The total energy calculated from the Heisenberg model is therefore invariant with respect to such a rotation, and the Goldstone condition is fulfilled identically. However, in the advanced methods based on TDDFT or MBPT, where the excitation energies define the pole structure of a two-particle propagator, the situation is much less transparent. We already know that the collective excitations arise from the nodes of the denominator of Eq. (44). (There is a very similar equation in TDDFT (Buczek et al. 2009; Lounis et al. 2010).) So, to be fulfilled, the Goldstone condition requires $KW$ to have an eigenvector with eigenvalue 1 in the limits $\mathbf{q} \rightarrow \mathbf{0}$ and $\omega \rightarrow 0$. (We omit the indices in this section for simplicity.) Any numerical inaccuracy will lead to a slight deviation of the respective eigenvalue from 1 and, as a consequence, to a violation of the Goldstone condition. Often, this problem is circumvented by using the Goldstone criterion to fix a free parameter of the numerical scheme, thus, making a virtue of necessity. This free parameter has been chosen to be the effective interaction (Kotani and van Schilfgaarde 2008; Karlsson and Aryasetiawan 2000), the exchange-correlation kernel (Lounis et al. 2010, 2011), the bare susceptibility (Rousseau et al. 2012), or a scaling factor for the screened Coulomb interaction (Şaşıoğlu et al. 2010; Friedrich et al. 2014). Another possibility is an a posteriori correction of the resulting susceptibility (Buczek 2009; Buczek et al. 2009, 2011). However, in our case, such a pragmatic approach seems inappropriate. There is no mistaking that $W$ is the RPA screened interaction Eqs. (13) and (14) and $K$ is the two-particle propagator Eq. (25). So, strictly speaking, there is no room or justification for introducing a free parameter. Therefore, we analyzed the problem in more detail in Müller et al. (2016). We argued that there is an inconsistency between the free propagator $G$ (and, hence, $K$) and the screened interaction $W$, and it is this inconsistency that is responsible for the Goldstone violation. In this chapter, we go a step further and present a mathematical proof that constructing the single-particle propagator from a self-consistent Coulomb-hole screened-exchange (COHSEX) self-energy (Hedin 1965, 1999) should revoke the inconsistency provided that a complete basis is used for the solution of the Bethe-Salpeter equation (Eq. (32)). However, the Wannier basis and, in particular, the on-site approximation do not fulfill the latter criterion so that a finite gap error must still be expected in our approach. Nevertheless, numerical results for the bulk $3d$ transition metals iron, cobalt, and nickel show that the Goldstone violation is substantially reduced if the propagator is self-consistently renormalized with the COHSEX self-energy. In practice, the application of the COHSEX self-energy is considerably more time-consuming than standard LSDA calculations. Therefore, we discuss a correction
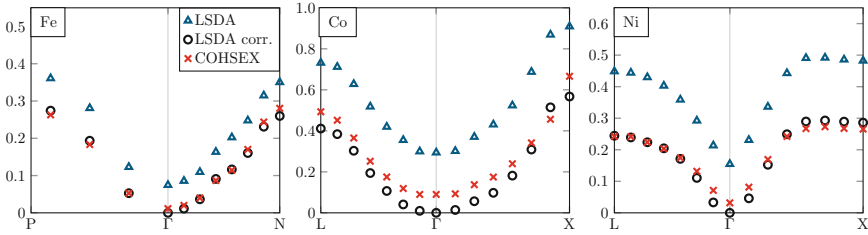
**Fig. 3** Spin-wave spectra for Fe, Co, and Ni obtained with the LSDA (blue triangles), corrected LSDA (black circles), and COHSEX (red crosses) Green functions as starting point. (Müller et al. (2016) Copyright 2016 American Physical Society)

scheme for the LSDA Green function of ferromagnetic materials, ultimately introducing a free parameter as a pragmatic and efficient solution. We show that the corrected LSDA magnon spectra for the $3d$ transition metals iron, cobalt, and nickel are close to the results obtained from the much more expensive COHSEX approach.

In the following, we discuss the spin-wave spectra for the elementary bulk ferromagnets Fe, Co, and Ni with regard to the starting-point dependence of MBPT. We refer here to the Green function used in Eqs. (14) and (25). Since a set of single-particle states is already available from the ground-state calculation, a convenient choice is the LSDA Green function calculated from the corresponding Kohn-Sham wave functions and energies. We have used Wannier functions of $s$, $p$, and $d$ character constructed from the 18 lowest Kohn-Sham bands. The resulting spin-wave dispersions for all three materials are shown as the blue symbols in Fig. 3, correctly showing a quadratic behavior around the $\Gamma$ point. However, they also clearly exhibit a violation of the Goldstone theorem: the spin-wave excitation energy does not vanish in the center of the BZ as it should.

There are a number of approximations used in our numerical approach, which might be responsible for this violation, e.g., the on-site approximation, the incompleteness of the Wannier basis, convergence issues (**k**-point set, basis sets, empty-state summations), and so on. Apart from these, there is another more fundamental inconsistency in the chosen approach, which we will investigate in the following. This inconsistency concerns the choice of the starting point, i.e., the LSDA Green-function propagator. Equation (44) is derived under the assumption that the Green function be self-consistently renormalized with the $GW$ self-energy. Only if this condition is fulfilled do we obtain the infinite series of ladder diagrams shown in Fig. 1. The two quantities, $G$ and $W$, are thus related, and one must be chosen in accordance with the other.

Unfortunately, fully self-consistent $GW$ calculations for transition-metal bulk systems are nowadays still a major challenge due to the dense **k**-point sets that are needed. On a second thought, however, we should also remember the static approximation that we have applied to $W$. For this reason, the proper self-energy to be used in the framework of our theoretical approach would have to be constructed

with the static screened interaction. An obvious choice would be the Coulomb-hole screened-exchange (COHSEX) self-energy (Hedin 1965, 1999), in which the dynamical $W$ is replaced by the static $W(\mathbf{r}, \mathbf{r}') = W(\mathbf{r}, \mathbf{r}'; \omega = 0)$. It consists of two terms, the screened exchange (SEX) and the Coulomb-hole (COH) term. The former corresponds to Hartree-Fock theory with the bare Coulomb interaction replaced by $W(\mathbf{r}, \mathbf{r}')$

$$\Sigma_{\text{SEX}}^{\sigma}(\mathbf{r}, \mathbf{r}') = -n^{\sigma}(\mathbf{r}, \mathbf{r}')W(\mathbf{r}, \mathbf{r}') \tag{49}$$

with the density matrix $n^{\sigma}(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{k}}^{\text{BZ}} \sum_{n}^{\text{occ}} \varphi_{\mathbf{k}n}^{\sigma}(\mathbf{r})\varphi_{\mathbf{k}n}^{\sigma*}(\mathbf{r}')$. The latter is given by

$$\Sigma_{\text{COH}}(\mathbf{r}, \mathbf{r}') = \frac{1}{2}\delta(\mathbf{r} - \mathbf{r}')[W(\mathbf{r}, \mathbf{r}') - v(\mathbf{r}, \mathbf{r}')], \tag{50}$$

which acts as a local and spin-independent potential. It accounts for the interaction energy of a quasiparticle with its induced (static) polarization cloud. Therefore, this term only couples charge degrees of freedom (if spin-orbit coupling is set aside) and does not affect the linear response of transversal spin fluctuations. Only Eq. (49), corresponding to Eq. (12) with $W(\mathbf{r}, \mathbf{r}'; \tau + \eta)$ replaced by $W(\mathbf{r}, \mathbf{r}')$, contributes to the right-hand side of Eq. (16) with $\delta \Sigma / \delta G = iW(0)$. Obviously, this leads to the same form of the Bethe-Salpeter equation as before.

Up to now, our argumentation was based on "theoretical consistency." In the following, we analyze the Goldstone criterion in a mathematical way starting from Eq. (44) (or the more general Eq. (32)), in which all quantities are four-point functions, and $-2R^{(4)}$ and $-2K$ must be understood in the general sense that they give the response of the magnetic density matrix $m^{+}(\mathbf{r}, \mathbf{r}'; \omega)$ with respect to changes of a nonlocal magnetic field $B^{+}(\mathbf{r}, \mathbf{r}'; \omega)$ in the interacting and noninteracting system, respectively. Conversely, the expression $(-2R^{(4)})^{-1}\Delta m^{+}$ gives the perturbing field $\Delta B^{+}(\mathbf{r}, \mathbf{r}'; \omega)$ that would generate the change of the magnetization $\Delta m^{+}(\mathbf{r}, \mathbf{r}; \omega)$. For the Goldstone mode in the limit $\omega \to 0$, we know that a rigid rotation of the electron spins, i.e., $\Delta m^{+} \propto m$, can take place even without a perturbing field. So, we have $(-2R^{(4)})^{-1}m = 0$, and with Eq. (44) we can write $KWm = m$. The eigenfunction of $KW$ with eigenvalue 1 is, thus, revealed to be the magnetization density (matrix). We claim that this condition is fulfilled if COHSEX is taken for the starting point. When separating off the spin-independent part of Eq. (49), the remaining spin-dependent part can formally be interpreted as a nonlocal magnetic field

$$B(\mathbf{r}, \mathbf{r}') = -\frac{1}{2}W(\mathbf{r}, \mathbf{r}')[n^{\uparrow}(\mathbf{r}, \mathbf{r}') - n^{\downarrow}(\mathbf{r}, \mathbf{r}')] = -\frac{1}{2}W(\mathbf{r}, \mathbf{r}')m(\mathbf{r}, \mathbf{r}'). \tag{51}$$

Now we use the simple fact that rigidly rotating the $\mathbf{B}$ field that creates the magnetization in a noninteracting system will rotate the magnetization in the same way, which can be expressed as

$$m(\mathbf{r}, \mathbf{r}') = -2 \iint K(\mathbf{r}, \mathbf{r}'; \mathbf{r}'', \mathbf{r}''') B(\mathbf{r}'', \mathbf{r}''') d\mathbf{r}'' d\mathbf{r}''' . \tag{52}$$

(For the rotation to be rigid, $\Delta B^+ \propto B$, and the corotation of $B$ field and magnetization requires $\Delta B^+/B = \Delta m^+/m$.) The assertion then follows from inserting Eq. (51) into Eq. (52).

The proof is already complete, but it is helpful to show $KWm = m$ more explicitly. In the limit $\omega \to 0$, the imaginary part of the two-particle propagator Eq. (34) is zero, and we can write

$$K(\mathbf{r}, \mathbf{r}'; \mathbf{r}'', \mathbf{r}''') = \frac{1}{N^2} \sum_{\mathbf{k}, \mathbf{k}'} \sum_{m, m'} (f_{\mathbf{k}m}^{\downarrow} - f_{\mathbf{k}'m'}^{\uparrow}) \frac{\varphi_{\mathbf{k}m}^{\downarrow}(\mathbf{r}) \varphi_{\mathbf{k}m}^{\downarrow *}(\mathbf{r}'') \varphi_{\mathbf{k}'m'}^{\uparrow *}(\mathbf{r}') \varphi_{\mathbf{k}'m'}^{\uparrow}(\mathbf{r}''')}{\varepsilon_{\mathbf{k}'m'}^{\uparrow} - \varepsilon_{\mathbf{k}m}^{\downarrow}} \tag{53}$$

with the occupation numbers $f_{\mathbf{k}m}^{\sigma}$. [We have used $\sum_m^{\mathrm{occ}} \sum_{m'}^{\mathrm{unocc}} \ldots = \sum_{m, m'} f_{\mathbf{k}m}^{\sigma} (1 - f_{\mathbf{k}'m'}^{\sigma}) \ldots = \sum_{m, m'} f_{\mathbf{k}m}^{\sigma} \ldots$ .] Because of Eq. (51) the COHSEX single-particle Hamiltonian fulfills $H^{\downarrow} - H^{\uparrow} = Wm$, which gives

$$\iint KWm = \frac{1}{N^2} \sum_{\mathbf{k}, \mathbf{k}'} \sum_{m, m'} (f_{\mathbf{k}m}^{\downarrow} - f_{\mathbf{k}'m'}^{\uparrow}) \frac{\varphi_{\mathbf{k}m}^{\downarrow}(\mathbf{r}) \langle \varphi_{\mathbf{k}m}^{\downarrow} | H^{\downarrow} - H^{\uparrow} | \varphi_{\mathbf{k}'m'}^{\uparrow} \rangle \varphi_{\mathbf{k}'m'}^{\uparrow *}(\mathbf{r}')}{\varepsilon_{\mathbf{k}'m'}^{\uparrow} - \varepsilon_{\mathbf{k}m}^{\downarrow}}$$

$$= \frac{1}{N} \sum_{\mathbf{k}} \sum_{m, m'} (f_{\mathbf{k}m'}^{\uparrow} - f_{\mathbf{k}m}^{\downarrow}) \varphi_{\mathbf{k}m}^{\downarrow}(\mathbf{r}) \langle \varphi_{\mathbf{k}m}^{\downarrow} | \varphi_{\mathbf{k}m'}^{\uparrow} \rangle \varphi_{\mathbf{k}m'}^{\uparrow *}(\mathbf{r}') = m(\mathbf{r}, \mathbf{r}') . \tag{54}$$

That the last expression is really the spin density matrix is seen by expanding $\varphi_{\mathbf{k}m'}^{\uparrow}(\mathbf{r})$ and $\varphi_{\mathbf{k}m}^{\downarrow *}(\mathbf{r})$ in terms of the functions of the other spin channel

$$m(\mathbf{r}, \mathbf{r}') = \frac{1}{N} \sum_{\mathbf{k}} \sum_{m'} f_{\mathbf{k}m'}^{\uparrow} \varphi_{\mathbf{k}m'}^{\uparrow}(\mathbf{r}) \varphi_{\mathbf{k}m'}^{\uparrow *}(\mathbf{r}') - \frac{1}{N} \sum_{\mathbf{k}} \sum_{m} f_{\mathbf{k}m}^{\downarrow} \varphi_{\mathbf{k}m}^{\downarrow}(\mathbf{r}) \varphi_{\mathbf{k}m}^{\downarrow *}(\mathbf{r}')$$

$$= \frac{1}{N} \sum_{\mathbf{k}} \sum_{m, m'} (f_{\mathbf{k}m'}^{\uparrow} - f_{\mathbf{k}m}^{\downarrow}) \varphi_{\mathbf{k}m}^{\downarrow}(\mathbf{r}) \varphi_{\mathbf{k}m'}^{\uparrow *}(\mathbf{r}') \langle \varphi_{\mathbf{k}m}^{\downarrow} | \varphi_{\mathbf{k}m'}^{\uparrow} \rangle . \tag{55}$$

In Müller et al. (2016), we reported on spin-wave calculations based on the COHSEX Green function. Technically, we start from the mean-field LSDA solution and construct the LSDA Green function, the corresponding polarization function Eq. (14), and the static screened interaction Eq. (13), from which the COHSEX self-energy Eqs. (49) and (50) is evaluated. The latter is a Hermitian operator defining a new mean-field system. This allows the respective single-particle equations of motion to be solved in a similar way as the Kohn-Sham equations of DFT. To be more precise, the single-particle equations are iteratively solved until the density is converged. This process updates the density and, consequently, the Hartree potential in each iteration, while the COHSEX self-energy matrix remains fixed. This
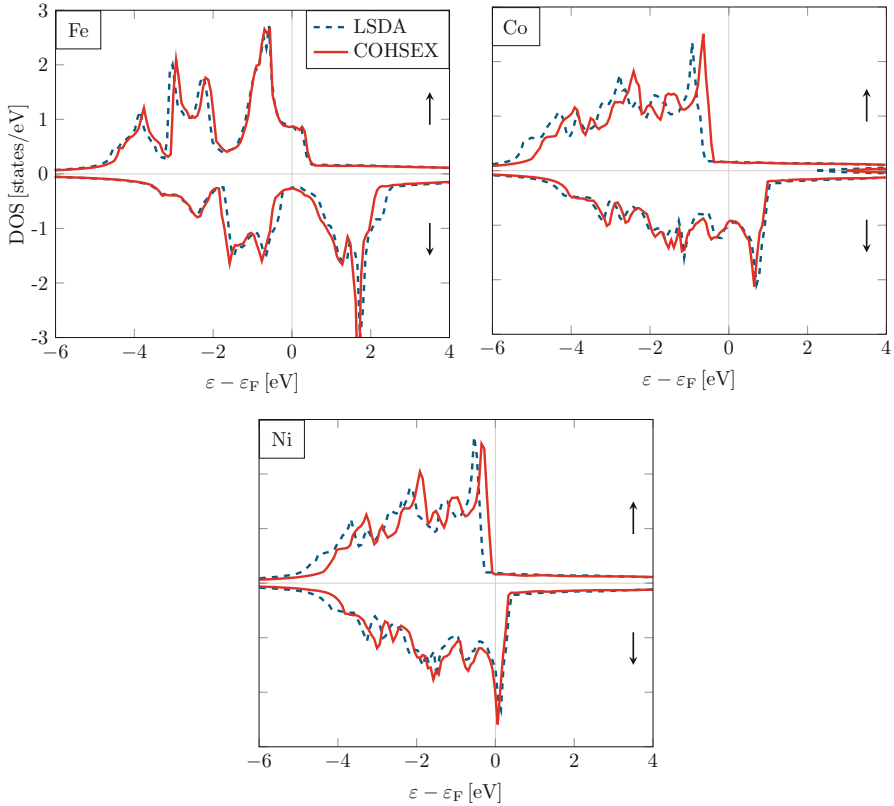
**Fig. 4** DOS spectra for bulk Fe, Co, and Ni. The Fermi level is set to zero. (Müller et al. (2016) Copyright 2016 American Physical Society)

produces a new set of wave functions and energies that are then used to construct a new Green function and, ultimately, a new COHSEX self-energy matrix. The whole procedure is repeated until self-consistency is achieved. We have employed a $14 \times 14 \times 14$ **k**-point set for these calculations.

The mean-field solution of the $3d$ ferromagnets bcc iron, fcc cobalt, and fcc nickel based on the COHSEX self-energy is interesting in its own right. Figure 4 shows their densities of states (DOS) for both LSDA and COHSEX. At a first glance, the two DOS spectra look very similar for all materials. The COHSEX self-energy yields thus qualitatively the same correct result as LSDA: all three materials are ferromagnetic metals. There are however small quantitative differences. The occupied bandwidth shrinks, in particular for Co and Ni, and the spin-up and spin-down states show a relative energetic shift toward each other. This observation is confirmed by the exchange splittings of selected single-particle states listed in Table 1. The COHSEX values are systematically smaller than the LSDA ones, to the effect that the slight overestimation of the magnetic moment found in LSDA is

**Table 1** Spin magnetic moments (obtained from a projection onto Wannier orbitals) and exchange splittings for selected states of Fe, Co, and Ni from LSDA, corrected LSDA, COHSEX, and experiment. (Müller et al. (2016) Copyright 2016 American Physical Society)

| | | | LSDA | LSDA corr. | COHSEX | Experiment |
|---|---|---|---|---|---|---|
| $m$ ($\mu_B$) | Fe | | 2.20 | 2.16 | 2.11 | 2.08 (Stearns 1986; Bonnenberg et al. 1986) |
| | Co | | 1.62 | 1.49 | 1.46 | 1.52 (Stearns 1986; Bonnenberg et al. 1986) |
| | Ni | | 0.59 | 0.51 | 0.46 | 0.52 (Stearns 1986; Bonnenberg et al. 1986) |
| $E_{ex}$ (eV) | Fe | $\Gamma'_{25}$ | 1.8 | 1.7 | 1.5 | 2.1 (Turner et al. 1984; Kisker et al. 1985; Sakisaka et al. 1985; Santoni and Himpsel 1991) |
| | | $H_{25}$ | 2.1 | 2.0 | 1.7 | 1.8 (Santoni and Himpsel 1991) |
| | | $P_4$ | 1.4 | 1.3 | 1.1 | 1.5 (Eastman et al. 1980) |
| | Co | $\Gamma'_{12}$ | 1.7 | 1.3 | 1.1 | 1.1 (Himpsel and Eastman 1980) |
| | | $\Gamma'_{25}$ | 1.4 | 1.0 | 1.2 | 1.1 (Himpsel and Eastman 1980) |
| | Ni | $L_3$ | 0.5 | 0.3 | 0.4 | 0.3 (Eastman et al. 1980) |
| | | $X_2$ | 0.6 | 0.4 | 0.3 | 0.2 (Raue et al. 1984) |

corrected to smaller values in COHSEX, albeit somewhat too strongly in the case of Co and Ni. With the exception of iron, the exchange splittings are improved by the self-consistent COHSEX calculation, most notably for Ni, whose exchange splitting is known to be overestimated in LSDA.

Figure 3 shows the spin-wave dispersion calculated from the COHSEX Green function as red symbols. Employing the self-consistent COHSEX mean-field solution as starting point, in fact, decreases the gap error systematically compared to the corresponding LSDA values. In case of bcc iron, fcc cobalt, and fcc nickel, the error is reduced by 85%, 69%, and 79%, respectively.

The ansatz presented so far is computationally very demanding. It requires the self-consistent calculation of the COHSEX self-energy on a fine **k**-point set. On the other hand, aside from the gap error, the magnon dispersions obtained from LSDA are very similar to the corresponding COHSEX results. This raises the question if it is possible to correct the LSDA Green function in a simple way that respects the Goldstone condition. In fact, this is possible.

Our approach is motivated by studying spin-wave solutions (Moriya 1985) of the one-band Hubbard model. When solved in the Hartree-Fock approximation, we obtain the magnetic susceptibility as a simple algebraic expression in the same form as Eq. (44) with the $W$ matrix replaced by the Hubbard interaction parameter $U$. In the Goldstone limit, the two-particle propagator simplifies to $K = m/E_{ex}$ with the site magnetization $m$ and the exchange splitting $E_{ex}$, and the Goldstone

condition can be phrased in the form of the simple relation $Um/E_{ex} = 1$. To remain consistent, we have to evaluate $K$ in the Hartree-Fock mean-field system, in which case $E_{ex} = Um$, and the Goldstone condition is identically fulfilled. The simple form of the relation invites one to use one of the constituent quantities as an adjustable parameter. The $U$ parameter plays the role of the screened interaction $W$, which is a matrix and thus cannot be corrected easily by a single parameter. Besides, we obtain $W$ from a many-body treatment of screening, and it does not seem appropriate to correct it in such an ad hoc way. Second, the magnetization $m$ results from the self-consistent LSDA calculation and cannot be varied straightforwardly. At last, $E_{ex}$ can be regarded as the energy difference between the spin-up and spin-down electron bands, which can easily be varied once a self-consistent LSDA solution has been found. Moreover, this correction will specifically modify the LSDA Green function, which was our intention, while leaving the screened interaction unchanged. The correction can be hoped to mimic to some extent the missing renormalization in the Green function. Therefore, we choose $E_{ex}$ as an adjustable parameter. To be more precise, we rigidly shift the spin-up and spin-down states relative to each other $\varepsilon_{\mathbf{k}m}^{\uparrow/\downarrow} \rightarrow \varepsilon_{\mathbf{k}m}^{\uparrow/\downarrow} \pm \Delta E_{ex}/2$ until the Goldstone condition is fulfilled. The LSDA Green function corrected in this way is then used to construct the two-particle propagator $K$. This procedure yields magnon dispersions, which respect the Goldstone condition and are close to the COHSEX results for the three materials as shown in Fig. 3. The relative shift in the band energies is such that the exchange splittings decrease. For Fe, Co, and Ni, we find $\Delta E_{ex} = 0.10$ eV, $\Delta E_{ex} = 0.39$ eV, and $\Delta E_{ex} = 0.21$ eV. The Fermi energy is adjusted accordingly so that the correction affects the ground-state magnetic properties as well. Interestingly, the resulting magnetic moments and exchange splittings turn out to be close to the corresponding COHSEX values listed in Table 1. They also compare well with experiment. The proximity of COHSEX and corrected LSDA values can be regarded as an a posteriori justification of the correction. Among the three materials, fcc cobalt appears as a problematic case. The gap error is largest and the COHSEX spin-wave dispersion shows an unusually flat behavior at the $\Gamma$ point. In fact, the curvature there is very small, being between results from LSDA (small positive curvature) and PBE (small negative curvature, not shown), indicative of a magnetic instability. This is in accordance with previous DFT results. Janak (1978) found that there are two competing magnetic ground states with low and high magnetic moment, and Moruzzi et al. (1986) report an unusually strong dependence of the magnetic properties on the lattice constant.

The findings can be interpreted in a more fundamental way. Formally, the Hamiltonian which describes the magnetic system is invariant with respect to spin rotations, while the ferromagnetic ground state is not. This implies the existence of a gapless excitation due to a homogeneous magnetic perturbation perpendicular to the magnetization axis. Baym and Kadanoff (1961) and Baym (1962) formulated a conserving and self-consistent scheme for correlation functions. The scheme was extended by Brandt et al. (Brandt et al. 1970, 1971; Brandt 1971) to the magnetic case. They showed that for a spin-conserving formulation of the magnetic susceptibility, which fulfills the Goldstone theorem automatically, several conditions

have to be fulfilled. The chosen self-energy approximation is to be calculated self-consistently with the Green function. This ensures that the single-particle states which form the basis for the electron-hole propagator are consistent with the applied self-energy approximation. In addition, the spin-independent interaction that is responsible for the correlation among the electron-hole pairs with opposite spins is required to be consistent with the self-energy as these properties are connected via $\delta\Sigma/\delta G = iW$. If both conditions are fulfilled, the magnetic response function will fulfill the Goldstone theorem. Then, the electron-hole pair propagator and the screened interaction are compatible with the Ward identity ensuring spin conservation. In particular, the correct limit $\mathbf{q} \to \mathbf{0}$ is attained.

# 5    Spin Excitation Spectra

In this section, we present detailed results of first-principles calculations for the three bulk transition metals Fe, Co, and Ni. The properties of these materials are strongly governed by the presence of the localized $3d$ states. It is the exchange interaction among the $3d$ electrons that drives the systems into the ferromagnetic ground state. On the other hand, the materials are metallic. The $3d$ states of neighboring atoms overlap, and there is a partly filled itinerant $4s$ band which spans the entire valence region and mixes with the $d$ bands. As a consequence, Fe, Co, and Ni show signatures of itinerant magnetism. For example, they fulfill the Stoner criterion of ferromagnetism, and there is no order-disorder phase transition at the Curie temperature as would be described by the Heisenberg model.

So, one would expect these bulk transition metals to show both localized and itinerant magnetic behavior. The present formalism using MBPT is capable of describing both types of magnetism on the same footing. The spectrum of spin excitations in this theory is given by the imaginary part of the magnetic response function as calculated from Eqs. (28) and (44). The poles of this function lie infinitesimally below (above) the positive (negative) real-frequency axis, and they come from both the two-particle propagator (bare susceptibility) in the numerator and the roots of the denominator. In the former case, the spin excitations have a single-particle character. These Stoner excitations can be described as excitations of single electrons across the Fermi surface with an accompanying spin flip of the electrons. In the latter case, the excitations are collective in nature, again with a total spin flip of 1, and arise from superpositions of infinitely many electron-hole pairs (single-particle excitations) coupled to each other by the exchange interaction. These electron-hole pairs describe changes in the spin density, in which all electrons take part collectively and which can, for example, have the form of spin waves. The two types of excitations are just limiting cases. In general, the spin excitations have a mixed character of single-particle and collective excitations: spin waves acquire a finite lifetime through a coupling to Stoner excitations, and Stoner excitations lose or gain spectral weight by a coupling to spin waves.

It is instructive to consider the model of a homogeneous electron gas with spin polarization. The spin-up and spin-down bands have the form of free-electron bands,

but they are shifted with respect to each other by the exchange splitting $E_{ex}$, i.e., $\varepsilon_{\mathbf{k}}^{\sigma} = k^2 - \sigma E_{ex}/2$. Energies and momenta are in units of the Fermi wavevector $k_0$ and Fermi energy $\varepsilon_0 = k_0^2/2$ of the paramagnetic system, respectively. The bare susceptibility of this system can be calculated from Eqs. (34) and (47) in a pure plane-wave representation. To simplify the evaluation of $K^{\downarrow\uparrow}$, we replace $-i\eta$ in the first term by $+i\eta$, which amounts to using the corresponding retarded quantity. This enables a straightforward integration over the occupied spin-up and spin-down band (Moriya 1985) yielding

$$
K^{+-}(q, \omega) = \frac{-1}{4\pi^2 q} \sum_{\sigma=-1}^{1}
$$

$$
\sigma \left[ \frac{1}{2}(p_\sigma^2 - k_\sigma^2) \left( \ln \left| \frac{p_\sigma + k_\sigma}{p_\sigma - k_\sigma} \right| + i\pi \, \mathrm{sgn}(\omega)\theta(k_\sigma - |p_\sigma|) \right) - p_\sigma k_\sigma \right]
\tag{56}
$$

with $p_\sigma = (\omega - \sigma q^2 - E_{ex})/(2q)$, the spin-dependent Fermi wavevector $k_\sigma = (1+\sigma\zeta)^{1/3}$, the exchange splitting $E_{ex} = (1+\zeta)^{2/3} - (1-\zeta)^{2/3}$, the spin polarization $\zeta = M/N = (k_\uparrow^3 - k_\downarrow^3)/(k_\uparrow^3 + k_\downarrow^3)$, and the Heaviside function $\theta(x) = 1$ for $x > 0$ and $0$ for $x < 0$. The factor $\mathrm{sgn}(\omega)$ has been introduced to undo the sign change $-i\eta \to +i\eta$, which recovers the time-ordered two-particle propagator. All quantities can be written in terms of a single parameter, the spin polarization $\zeta \in [0, 1]$.

   The area where the imaginary part of $K^{+-}(q, \omega)$ is nonzero defines the Stoner continuum, which, by using the definition of $\theta$, can be shown to be bounded by the functions $2qk_\sigma + \sigma q^2 + E_{ex}$ from above and $-2qk_\sigma + \sigma q^2 + E_{ex}$ from below. Figure 5 presents a plot of $-\mathrm{Im}\, K^{+-}(\mathbf{q}, \omega)$ together with the boundary lines for a spin-polarized electron gas with $\zeta = 0.9$. The imaginary part diverges for $\omega = E_{ex} = 1.32$ and $q \to 0$, because the real part of the denominator of Eq. (34) vanishes in this limit. We also see that the Stoner continuum extends toward negative energies. This is because, as long as $\zeta < 1$, there can be transitions from occupied minority to unoccupied majority bands. At $\omega = 0$, the imaginary part of $K^{+-}$ is zero for all momenta because the phase space of single-particle excitations vanishes in this limit.

   Of course, the homogeneous electron gas is a relatively crude model. The absence of the crystal field and of atomic wave functions makes one wonder whether this model could be just too simple. Therefore, it is interesting to compare Fig. 5a to corresponding plots calculated with realistic wave functions and energies from a self-consistent Kohn-Sham solution. In particular, we have calculated the bare susceptibility from Eq. (48) and projected it onto the plane wave $e^{i\mathbf{q}\cdot\mathbf{r}}$ from both sides giving $K^{+-}(\mathbf{q}, \omega)$. A fine $20 \times 20 \times 20$ $\mathbf{k}$-point grid was employed, and the Wannier basis was the same as in Sect. 4. We have also applied the $E_{ex}$ correction. The plots for bcc Fe, fcc Co, and fcc Ni are presented in Fig. 6a–c. It is surprising that they share a number of similarities with the plot for the homogeneous electron gas. First, the maximum of $-\mathrm{Im}\, K^{+-}(\mathbf{q}, \omega)$ is seen for $\omega \approx E_{ex}$ and $\mathbf{q} = \mathbf{0}$. In
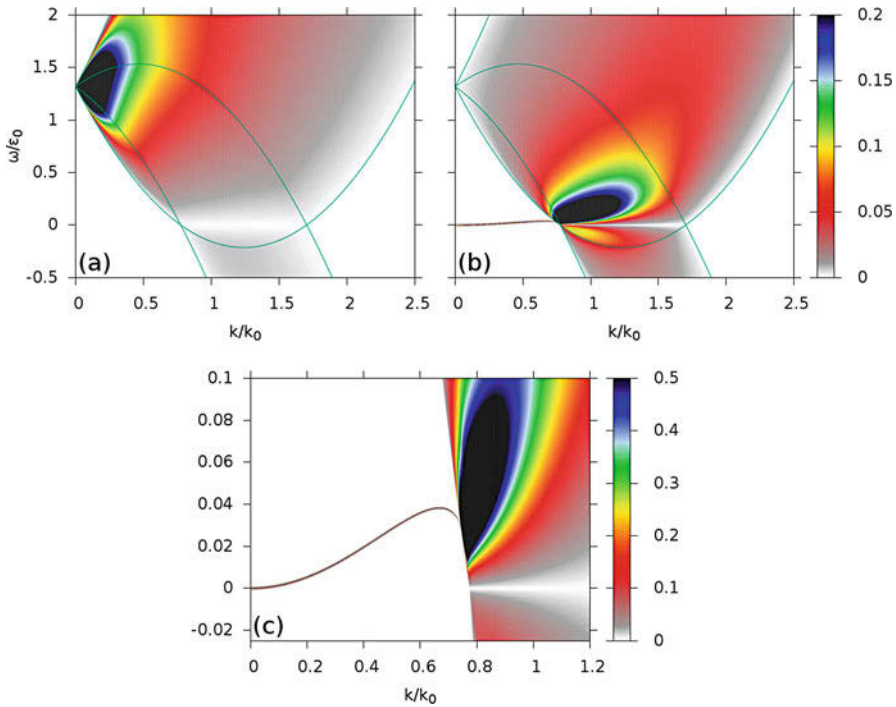
**Fig. 5** Imaginary part of (**a**) the bare $[-\operatorname{Im} K^{+-}(\mathbf{q}, \omega)]$ and (**b**) the renormalized susceptibility $[-\operatorname{Im} R^{+-}(\mathbf{q}, \omega)]$ for the homogeneous electron gas with spin polarization $\zeta = 0.9$; (**c**) is a magnification of (**b**). The green solid lines show the boundaries of the Stoner continuum. In (**b**) and (**c**), a finite imaginary frequency of, respectively, $i\eta = i10^{-5}$ and $i\eta = i10^{-7}$ has been employed in order to make the magnon branch visible, which would have a vanishing (delta-like) width in this system otherwise

contrast to before, there is no divergence because spin-up and spin-down bands have different dispersions in a real material, and the majority bands shifted up by $E_{ex}$ thus cannot coincide exactly with the minority bands, in particular, since $E_{ex}$ is not a unique quantity but $\mathbf{k}$ dependent. Consequently, the maximum at $\mathbf{q} = \mathbf{0}$ has a certain width in energy. Furthermore, as before, we observe a weak intensity of Stoner excitations for negative $\omega$. As a qualitative difference to Fig. 5, the spectra of Co and Ni exhibit a horizontal, nearly dispersion-less band of high intensity that emanates from the maximum at $\Gamma$ and stretches toward the X point. This feature can be directly related to the localized nature of the single-particle states. The densities of $d$ states of Co and Ni show particularly sharp peaks in the majority valence and minority conduction regions separated by 1.3 and 0.5 eV, respectively, revealing the feature to originate from single-particle $d \rightarrow d$ transitions.

There are no clear boundary lines as for the homogeneous electron gas beyond which the imaginary part of $K^{+-}$ would vanish. In fact, there can be Stoner excitations for all $\mathbf{q}$ and $\omega$ (except for $\omega = 0$). However, the form of $-\operatorname{Im} K^{+-}(\mathbf{q}, \omega)$ is
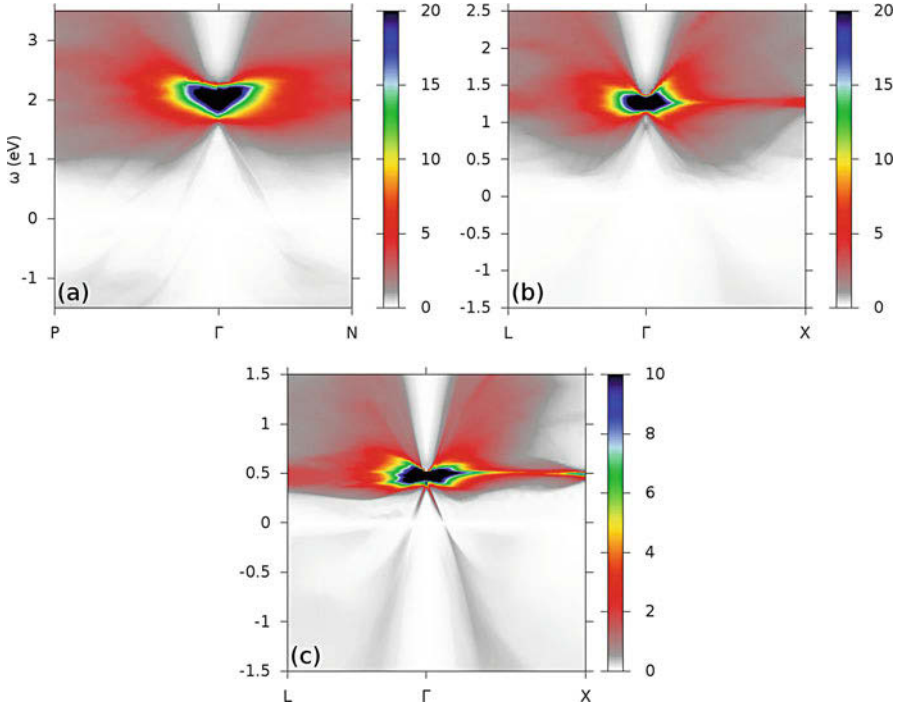
**Fig. 6** Imaginary part of the bare susceptibility $[-\operatorname{Im} K^{+-}(\mathbf{q}, \omega)$ in units of $(\Omega \mathrm{eV})^{-1}; \Omega = $ unit-cell volume] for (**a**) bcc Fe, (**b**) fcc Co, and (**c**) fcc Ni as a function of $\mathbf{q}$ and $\omega$

definitely similar. In particular, the characteristic V-shaped regions with (nearly) no excitations above and below the maximum are present, and, instead of the boundaries, one observes *stripes* of increased intensity seemingly extending from the maximum. A pronounced one is seen in the spectrum of Ni. These stripes will play an important role in the renormalized spectra later on. They can be understood by realizing that the function $-\operatorname{Im} K^{+-}(q, \omega)$ can roughly be thought of as giving the intersection of the valence majority states shifted upward by $\omega$ and sideways by $-\mathbf{q}$ with the conduction minority states. This is illustrated in Fig. 7 where the electronic bands are assumed to show a linear dispersion at finite $q$. Starting from the (near) coincidence of the bands at $q = 0$ and $\omega = E_{\mathrm{ex}}$, relatively large intensities are still expected when the majority band is shifted from there in such a way that the regions of linear dispersion remain overlapping, giving rise to the condition of proportionality $\omega \propto q$ and, hence, to the stripes. Here, we have assumed that the spin-up band is filled, while the spin-down band is empty and that they exhibit similar dispersions as is often the case in ferromagnets. Even in the case of the highly symmetric homogeneous electron gas, whose bands exhibit a constant curvature and no linear dispersions as in Fig. 7, the onset of the Stoner continuum at the boundaries is quite abrupt. So, it is not surprising that real materials show more structure there.

**Fig. 7** Illustration of the "band overlap" leading to the striped Stoner continuum of Fig. 6. The red arrow denotes the proportionality $\omega \propto q$
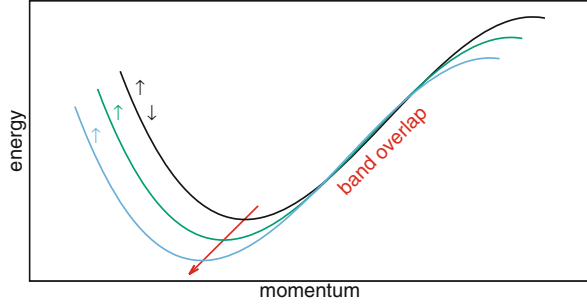
Up to now, we have discussed the spin excitation spectrum of the fictitious noninteracting reference system. The spectrum changes profoundly when one introduces a finite interaction among the electrons. In the case of the homogeneous electron gas, one would have to solve the Bethe-Salpeter equation $R^{+-}(q,\omega) = K^{+-}(q,\omega)/[1 - WK^{\downarrow\uparrow}(q,\omega)]$ with $K^{\downarrow\uparrow}(q,\omega) = -K^{+-}(q,\omega)/2$ (cmp. Eq. (47)) and a so-far unknown effective interaction $W$. The latter is a single parameter in this case and can be obtained conveniently from the Goldstone condition, giving $W = 1/K^{\downarrow\uparrow}(0,0) = -2/K^{+-}(0,0) = 3\pi^2 E_{ex}/\zeta$. The spin excitation spectrum is then given by $-\mathrm{Im}\,R^{+-}(\mathbf{q},\omega)$ shown in Fig. 5b for the homogeneous electron gas. Together with the spectrum, we have plotted the boundary lines of the Stoner continuum as solid green lines. Still, these lines separate the regions of finite intensity from the regions of no intensity, except for the spin-wave branch, which starts at the origin $q = \omega = 0$ and disperses quadratically for small $q$ according to $\omega \propto Dq^2$ with $D = [1 - 0.4(k_\uparrow^5 - k_\downarrow^5)/(k_\uparrow^2 - k_\downarrow^2)]/\zeta$ (Moriya 1985). For larger $q$, the magnon branch deviates from the parabolic dispersion and finally enters the Stoner continuum, where it couples to the Stoner excitations forming a broad maximum. The magnon energies are much smaller than $E_{ex}$, even for small $E_{ex}$ because $D \sim E_{ex}/12$ in this limit. Therefore, we show an enlarged picture of the magnon branch in Fig. 5c. We note that a finite (instead of infinitesimal) parameter $\eta = 10^{-5}$ and $\eta = 10^{-7}$ has been employed in (b) and (c), respectively, which leads to a corresponding finite linewidth of the magnon branch outside the Stoner continuum.

When comparing with Fig. 5a, we observe a strong redistribution of quasiparticle weight after solving the Bethe-Salpeter equation. In particular, the region of maximal intensity around $q = 0$ now appears very shallow, and a new maximum is found where the spin-wave branch enters the Stoner continuum. The former is a feedback effect: the transfer of spin-up electrons into the spin-down channel leads to a strong change in the exchange field of the interacting system, which acts against the transfer of electrons, an effect similar to the electronic screening effect. The new maximum comes from a resonance effect between the collective magnon and the Stoner excitations. The amplifying effect of the resonance extends into the Stoner excitations of negative energy ($\omega < 0$), which now appear more intense than in Fig. 5a. It is interesting to note that the maximum in Fig. 5a transforms continuously into the one of (b) if one smoothly varies the parameter $W$ from 0 to $3\pi^2 E_{ex}/\zeta$.

The spin-wave branch starts to appear for $W \gtrsim 0$ slightly below the maximum at $\omega = E_{ex}$, and as $W$ increases, it shifts down, first with a maximum at $q = 0$, then developing a minimum, until, for $W = 3\pi^2 E_{ex}/\zeta$, it finally has the form of Fig. 5c. Further increasing $W$ would shift the minimum of the magnon branch to negative energies. This again demonstrates how sensitive the Goldstone condition is and explains the difficulty of its fulfillment in first-principles calculations.

In Fig. 8a–c, we show the renormalized susceptibility for Fe, Co, and Ni, again projected onto $e^{i\mathbf{q}\cdot\mathbf{r}}$. We have used the same scales as before to make a direct comparison possible. Obviously, the regions around the spin-wave branch have values exceeding the maximum of the color scale (black area). Therefore, Fig. 8d–f presents a magnified picture with a different color scale. As with the bare susceptibility, there are a number of similarities to Fig. 5b: a spin-wave branch is formed by the renormalization; this branch acquires a finite lifetime broadening by a coupling to Stoner excitations; the former intensity maximum around $\omega = E_{ex}$ and $\mathbf{q} = \mathbf{0}$ has lost much of its intensity; there is a resonant enhancement of the Stoner intensity for $\omega < 0$. However, there are also some important differences to the case of the homogeneous electron gas. First, the coupling to Stoner excitations sets in much earlier, because the Stoner excitations are present at all energies. Nevertheless, the spin-wave dispersion remains discernible to quite large momenta and energies compared to the case of the homogeneous electron gas, indicating a localized nature of magnetism in these materials. Especially in Ni, the magnon branch seems to couple resonantly with the horizontal Stoner band toward the X point. While the spin-wave branches still have an overall parabolic shape, they show a rather irregular behavior, which can be attributed to the coupling to the Stoner continuum. In particular, in all cases the magnon branch is affected, sometimes interrupted, by the interaction with the *striped* Stoner spectrum we have discussed before. In fact, if we magnify the corresponding region of the spectrum for Ni (see inset in Fig. 8f), we observe that the coupling between the magnon branch and a line of strong Stoner intensity – we could call it a *Stoner band* – leads to a feature that looks like an "avoided crossing." In fact, this feature has a similar origin as an avoided crossing of single-particle bands, with the difference that the two states that interact here are not single-particle but many-body states, the collective spin-wave excited state and a superposition of Stoner excitations, which mix and exchange character.

Despite the multiple interactions with the Stoner background, we find that the dispersion of the spin-wave branches is mostly isotropic in $\mathbf{q}$ space. The numerical data for iron are in qualitative and also quantitative agreement with the neutron scattering experiments of Collins et al. (1969), Mook and Nicklow (1973), and Lynn (1975), where spin-wave energies up to 70, 118, and 110 meV were reported, respectively; see Friedrich et al. (2014) for a comparison. In Loong et al. (1984), spin-wave resonances up to an energy of 160 meV could be measured. However, the latter experimental results partly disagree with the values of Mook and Nicklow (1973), especially at the high end of the spectrum. We also find good agreement in the case of nickel (Minkiewicz et al. 1969; Mook and Tocchetti 1979; Mook and Paul 1985), while neutron scattering data for cobalt is scarce and limited to small momentum transfers (Frikkee 1966; Glinka et al. 1977).
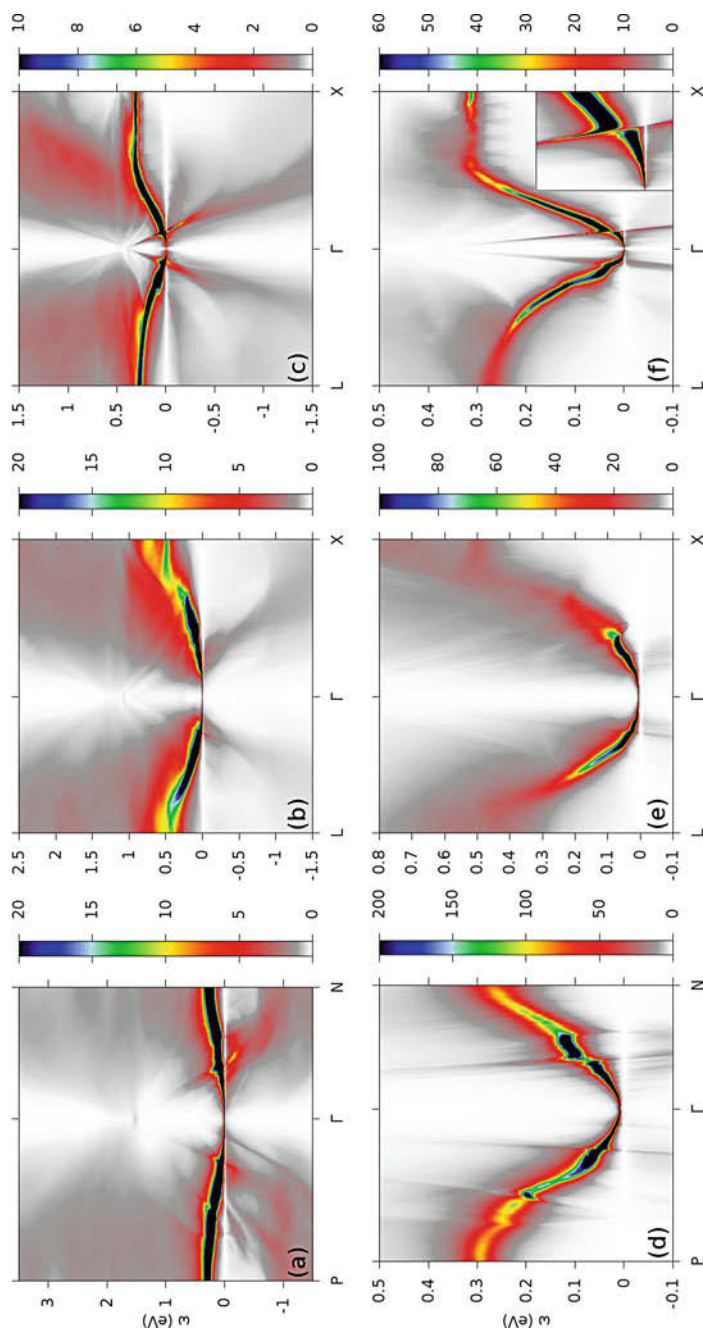
**Fig. 8** Same as Fig. 6 for the renormalized susceptibility $[-\operatorname{Im} R^{+-}(\mathbf{q}, \omega)$ in units of $(\Omega\, eV)^{-1}]$; (**d**)–(**f**) are magnifications of (**a**)–(**c**) with different scales. The inset of (**f**) shows a magnified view of the region where the magnon branch couples to the increased Stoner intensity

A feature that has been discussed extensively in the literature (Cooke 1976; Mook and Tocchetti 1979; Cooke et al. 1980, 1985; Callaway et al. 1983; Mook and Paul 1985; Savrasov 1998; Karlsson and Aryasetiawan 2000; Şaşıoğlu et al. 2010; Friedrich et al. 2010) is the appearance of an "optical" spin-wave branch, in addition to the acoustic one, for example, in fcc Ni along $\Gamma - X$. The terminology is, however, a little bit misleading, as the acoustic branch just seems to exhibit a gap at around 130 meV. The lower-energy end flattens and eventually disappears toward larger momenta (or persists out to the zone boundary (Cooke et al. 1985; Blackman et al. 1985)), and the higher-energy end then continues to form the acoustic magnon branch. The two ends coexist in a certain region of the reciprocal space giving a double-peak structure in the spectra there. This indicates the existence of an optical branch that crosses the acoustic branch and splits it into two pieces. The optical branch has yet to be observed directly, however. (In some publications, the higher-energy branch is denoted as the optical branch, although it does not extend to zero momentum and finite energy as in the case of optical phonon branches.) After its theoretical prediction (Cooke 1976), the optical branch was detected in the form of a double peak in constant $\mathbf{q}$ scans in neutron scattering experiments (Mook and Tocchetti 1979; Mook and Paul 1985). However, this feature seems to be an elusive phenomenon, which is observed in some studies (Cooke et al. 1985; Blackman et al. 1985) but not in others (Callaway et al. 1983).

In Fig. 8f, we clearly see a gap in the acoustic branch at around 30 meV, which arises from the coupling to Stoner excitations as we have seen before. But this gap is located at a too low energy and too close to the $\Gamma$ point to be identified as the optical branch discussed in the literature. On the other hand, there is no gap to be seen in Fig. 8f at around 130 meV. However, if we analyze this part of the excitation spectrum more closely, we are able to identify two peaks, but these peaks do not appear as separate peaks in our calculation. They form a broad peak, and only a peak fit with Lorentzian functions reveals the existence of a lower-energy and a high-energy branch in the respective region of momentum (Şaşıoğlu et al. 2010). In the direction of increasing momentum, the higher-energy peak grows at the expense of the lower-energy peak, which eventually disappears, and the higher-energy peak then forms the magnon branch. It should be noted, however, that the double-peak structure was observed as two separate peaks in a TDDFT study (Savrasov 1998) and also in a calculation based on a Green-function formulation (Karlsson and Aryasetiawan 2000). The latter study also reported a gapped magnon branch along $\Gamma - L$ in fcc Ni and a gap around halfway on the line $\Gamma - N$ in bcc Fe. In fact, there is a gap in this direction in Fig. 8d, albeit at a smaller momentum and energy than in Karlsson and Aryasetiawan (2000). This gap can be attributed to the coupling of the spin-wave branch with a line of increased Stoner intensity. The gapped magnon branch has already been discussed by Blackman et al. (1985) and was observed experimentally (Paul et al. 1988). In conclusion, it remains an open question why the optical branch in fcc Ni appears as a well-defined feature in some calculations, whereas, in others, it is so close to the acoustic branch that the two branches coalesce into a broad peak.

# 6     Conclusions

We have presented a theoretical method to determine the electronic spin excitations of an interacting many-electron system from first principles. The scheme is based on many-body perturbation theory, in which the spin excitations form the pole structure of the magnetic response function or transverse spin susceptibility. The poles are close to the real-frequency axis, so the imaginary part of the response function yields the corresponding spin excitation spectrum, comprising both single-particle Stoner and collective spin-wave excitations as well as combinations thereof. The latter gives rise to lifetime effects and a redistribution of spectral weight. We have described a very general theoretical derivation, in which the density response function, which is central in describing optical absorption and excitonic effects, appears as a special case.

With the $GW$ approximation for the electronic self-energy, the magnetic response function has been shown to fulfill a Bethe-Salpeter equation, which can be solved in the basis of Wannier product functions. We have sketched a practical implementation in the SPEX code that relies on the full-potential linearized augmented plane-wave method. The screened interaction $W$ is calculated within the RPA. In metallic systems, $W$ falls off very quickly so that we can afford to employ an on-site approximation, i.e., an electron-hole pair is assumed to be on the same site when interacting. In addition, we use the static limit of $W$ for all frequencies, so the screened interaction acts instantaneously. Our implementation allows the magnetic response function to be calculated for arbitrary momenta, which can be used to map the magnetic excitation spectra in very fine detail.

We have studied the long-wavelength limit of the spin-wave spectra for the bulk $3d$ transition metals Fe, Co, and Ni. The long-wavelength limit is of special interest as the Goldstone theorem demands the existence of a gapless excitation in ferromagnetic materials (neglecting spin anisotropy). Often, this *Goldstone condition* is numerically violated in practical calculations from first principles. We can attribute a large part of this gap error to the approximation of the single-particle Green function. For example, a natural and convenient choice would be the LSDA Green function, which, however, introduces an inconsistency with regard to the theoretical derivation of the Bethe-Salpeter equation. We have proved mathematically that a gapless magnon branch requires the Green function to be self-consistently renormalized with a suitable self-energy, e.g., the COHSEX self-energy, the static limit of the $GW$ approximation. It was shown numerically that the gap error is substantially reduced when using the COHSEX Green function instead of the LSDA one. Furthermore, the self-consistent COHSEX calculations give rise to an overall reduction of the exchange splitting compared to LSDA, often leading to better agreement with experiment. The spin-wave solution of the one-band Hubbard model employing the Hartree-Fock approximation motivates a correction scheme for the LSDA Green function, where the exchange splitting of the Kohn-Sham system is adjusted so as to enforce the Goldstone condition. The resulting spin-wave dispersions are closer to the corresponding COHSEX than to the original

LSDA results. The same can be said about the magnetic moments and exchange splittings obtained from the COHSEX and the corrected LSDA Green function, which are found to be very similar, while the original LSDA values are a bit off. As a result, the corrected LSDA Green function mimics that of the self-consistent COHSEX calculation and is made to fulfill the Goldstone condition exactly, while the numerical cost is identical to a treatment within LSDA. This opens up the possibility of efficient first-principles MBPT calculations of spin excitations that respect the Ward identity of spin conservation.

Spin excitation spectra of the three elementary ferromagnets bcc Fe, fcc Co, and fcc Ni have been calculated and compared to results of model calculations using a spin-polarized homogeneous electron gas. By this comparison, we have shown that the bulk ferromagnets exhibit many features of itinerant-electron magnets that cannot be described by a simple atomic arrangement of magnetic moments, such as in the Heisenberg model. High-energy magnons are strongly damped due to the coupling to single-particle Stoner excitations, and the spin-wave dispersion possesses gaps along certain directions in the Brillouin zone, which can be attributed to a coupling of the spin-wave branch with *Stoner bands*, lines of increased intensity of Stoner excitations.

This coupling effect can be made responsible for the appearance of an *optical branch*, in addition to the acoustic one, in bcc iron. On the other hand, we cannot unambiguously identify an optical branch in fcc nickel along $\Gamma - X$, which has been much discussed in the literature. While a peak fit of the calculated spin excitation spectrum does reveal two spin-wave peaks at about the right momentum and energy, this appears as a very subtle effect compared to the gaps discussed before. Overall, we find a good agreement of the spin-wave dispersions to neutron scattering experiments.

The present treatment of spin excitations within many-body perturbation theory explicitly describes the correlated motion of an electron-hole pair. This formulation can be straightforwardly extended to yield the dynamical longitudinal spin susceptibility, including its coupling to the charge susceptibility (density response function). Furthermore, it opens up the way for constructing a diagrammatic electronic self-energy that describes the scattering of electrons and holes with magnons, in a similar way as the $GW$ approximation describes the scattering with plasmons. Such a self-energy can be defined by the product of the Green function with the $T$ matrix, yielding the $GT$ self-energy (Hertz and Edwards 1973; Edwards and Hertz 1973). A numerical implementation (Müller 2016; Müller et al. 2019) would be a formidable task given that the $T$ matrix depends on four points in space. It may be possible to combine this self-energy with $GW$ yielding $GW\Gamma$, a self-energy with vertex corrections. Such a self-energy could be able to describe the quasiparticle renormalization (kinks) of electronic bands in magnetic materials due to the scattering with spin fluctuations. It might even shed light on the coupling mechanism in high-temperature superconductors, as it is believed that this coupling arises from the exchange of magnons, through which the effective electron-electron interaction can become attractive.

# References

Aryasetiawan F, Karlsson K (1999) Green's function formalism for calculating spin-wave spectra. Phys Rev B 60:7419–7428. https://doi.org/10.1103/PhysRevB.60.7419

Baym G (1962) Self-consistent approximations in many-body systems. Phys Rev 127(4):1391–1401. https://doi.org/10.1103/PhysRev.127.1391

Baym G, Kadanoff LP (1961) Conservation laws and correlation functions. Phys Rev 124(2):287–299. https://doi.org/10.1103/PhysRev.124.287

Blackman JA, Morgan T, Cooke JF (1985) Prediction of high-energy spin-wave excitation in iron. Phys Rev Lett 55:2814–2817. https://doi.org/10.1103/PhysRevLett.55.2814

Bloch F (1930) Zur theorie des ferromagnetismus. Zeitschrift für Physik 61(3):206–219. https://doi.org/10.1007/BF01339661

Bonnenberg D, Hempel K, Wijn H (1986) 3D, 4D and 5D elements, alloys and compounds. In: Wijn HP, Landolt H, Börnstein R (eds) Magnetic properties in metals, new series, vol III. Springer, Berlin. https://doi.org/10.1007/b29710, http://materials.springer.com/bp/docs/978-13-540-39667-3?utm_campaign=bookshelf-experiment&utm_medium=xls&utm_source=staticpage

Brandt U (1971) Modified T-matrix approximation in itinerant ferromagnets. Zeitschrift für Physik 244(3):217–229. https://doi.org/10.1007/BF01395567

Brandt U, Pesch W, Tewordt L (1970) Self-consistent approximations for itinerant ferromagnets above the phase transition point. Zeitschrift für Physik 238(2):121–129. https://doi.org/10.1007/BF01399298

Brandt U, Lustfeld H, Pesch W, Tewordt L (1971) Self-consistent approximations for itinerant ferromagnetism below the phase-transition point. J Low Temp Phys 4(1):79–95. https://doi.org/10.1007/BF00628439

Buczek P (2009) Spin dynamics of complex itinerant magnets. Ph.D. thesis, Martin-Luther-Universität Halle Wittenberg

Buczek P, Ernst A, Bruno P, Sandratskii LM (2009) Energies and lifetimes of magnons in complex ferromagnets: a first-principle study of Heusler alloys. Phys Rev Lett 102:247206. https://doi.org/10.1103/PhysRevLett.102.247206

Buczek P, Ernst A, Sandratskii LM (2010) Standing spin waves as a basis for the control of terahertz spin dynamics: time dependent density functional theory study. Phys Rev Lett 105:097205. https://doi.org/10.1103/PhysRevLett.105.097205

Buczek P, Ernst A, Sandratskii LM (2011) Interface electronic complexes and landau damping of magnons in ultrathin magnets. Phys Rev Lett 106:157204. https://doi.org/10.1103/PhysRevLett.106.157204

Callaway J, Chatterjee AK, Singhal SP, Ziegler A (1983) Magnetic susceptibility of ferromagnetic metals: application to nickel. Phys Rev B 28:3818

Collins MF, Minkiewicz VJ, Nathans R, Passell L, Shirane G (1969) Critical and spin-wave scattering of neutrons from iron. Phys Rev 179:417–430. https://doi.org/10.1103/PhysRev.179.417

Cooke JF (1973) Neutron scattering from itinerant-electron ferromagnets. Phys Rev B 7:1108–1116. https://doi.org/10.1103/PhysRevB.7.1108

Cooke JF (1976) Magnetic excitations in itinerant electron systems. In: Moon RM (ed), Proceedings of the conference on neutron scattering. NTIS, Springfield, VA, 2:723

Cooke JF, Lynn JW, Davis HL (1980) Calculations of the dynamic susceptibility of nickel and iron. Phys Rev B 21:4118–4131. https://doi.org/10.1103/PhysRevB.21.4118

Cooke JF, Blackman JA, Morgan T (1985) New interpretation of spin-wave behavior in nickel. Phys Rev Lett 54:718–721. https://doi.org/10.1103/PhysRevLett.54.718

Dagotto E (1994) Correlated electrons in high-temperature superconductors. Rev Mod Phys 66:763–840. https://doi.org/10.1103/RevModPhys.66.763

Doniach S, Engelsberg S (1966) Low-temperature properties of nearly ferromagnetic Fermi liquids. Phys Rev Lett 17:750–753. https://doi.org/10.1103/PhysRevLett.17.750

Eastman DE, Himpsel FJ, Knapp JA (1980) Experimental exchange-split energy-band dispersions for Fe, Co, and Ni. Phys Rev Lett 44:95–98. https://doi.org/10.1103/PhysRevLett.44.95

Edwards DM, Hertz JA (1973) Electron-magnon interactions in itinerant ferromagnetism. II. Strong ferromagnetism. J Phys F Metal Phys 3(12):2191. http://stacks.iop.org/0305-4608/3/i=12/a=019

Faleev SV, van Schilfgaarde M, Kotani T (2004) All-electron self-consistent $GW$ approximation: application to Si, MnO, and NiO. Phys Rev Lett 93:126406. https://doi.org/10.1103/PhysRevLett.93.126406

Freimuth F, Mokrousov Y, Wortmann D, Heinze S, Blügel S (2008) Maximally localized Wannier functions within the FLAPW formalism. Phys Rev B 78:035120. https://doi.org/10.1103/PhysRevB.78.035120

Friedrich C, Blugel S, Schindlmayr A (2009) Efficient calculation of the coulomb matrix and its expansion around k = 0 within the FLAPW method. Comput Phys Commun 180:347

Friedrich C, Blügel S, Schindlmayr A (2010) Efficient implementation of the $GW$ approximation within the all-electron FLAPW method. Phys Rev B 81:125102. https://doi.org/10.1103/PhysRevB.81.125102

Friedrich C, Şaşıoğlu E, Müller M, Schindlmayr A, Blügel S (2014) Spin excitations in solids from many-body perturbation theory. In: Di Valentin C, Botti S, Cococcioni M (eds) First principles approaches to spectroscopic properties of complex materials, topics in current chemistry, vol 347. Springer, Berlin/Heidelberg, pp 259–301. https://doi.org/10.1007/128_2013_518

Frikkee E (1966) Inelastic scattering of neutrons by spin waves in F.C.C. cobalt. Physica 32(11):2149–2160

Glinka CJ, Minkiewicz VJ, Passell L (1977) Small-angle critical neutron scattering from cobalt. Phys Rev B 16:4084–4103. https://doi.org/10.1103/PhysRevB.16.4084

Halilov SV, Perlov AY, Oppeneer PM, Eschrig H (1997) Magnon spectrum and related finite-temperature magnetic properties: a first-principle approach. EPL (Europhys Lett) 39(1):91. http://stacks.iop.org/0295-5075/39/i=1/a=091

Hedin L (1965) New method for calculating the one-particle green's function with application to the electron-gas problem. Phys Rev 139:A796–A823. https://doi.org/10.1103/PhysRev.139.A796

Hedin L (1999) On correlation effects in electron spectroscopies and the GW approximation. J Phys Condens Matter 11(42):R489. http://stacks.iop.org/0953-8984/11/i=42/a=201

Hertz JA, Edwards DM (1973) Electron-magnon interactions in itinerant ferromagnetism. I. Formal theory. J Phys F Metal Phys 3(12):2174. http://stacks.iop.org/0305-4608/3/i=12/a=018

Himpsel FJ, Eastman DE (1980) Experimental energy-band dispersions and magnetic exchange splitting for cobalt. Phys Rev B 21:3207–3213. https://doi.org/10.1103/PhysRevB.21.3207

Hofmann A, Cui XY, Schäfer J, Meyer S, Höpfner P, Blumenstein C, Paul M, Patthey L, Rotenberg E, Bünemann J, Gebhard F, Ohm T, Weber W, Claessen R (2009) Renormalization of bulk magnetic electron states at high binding energies. Phys Rev Lett 102:187204. https://doi.org/10.1103/PhysRevLett.102.187204

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871. https://doi.org/10.1103/PhysRev.136.B864

Hong J, Mills DL (1999) Theory of the spin dependence of the inelastic mean free path of electrons in ferromagnetic metals: a model study. Phys Rev B 59:13840–13848. https://doi.org/10.1103/PhysRevB.59.13840

Hong J, Mills DL (2000) Spin dependence of the inelastic electron mean free path in Fe and Ni: explicit calculations and implications. Phys Rev B 62:5589–5600. https://doi.org/10.1103/PhysRevB.62.5589

Janak J (1978) Itinerant ferromagnetism in fcc cobalt. Solid State Commun 25(2):53–55. https://doi.org/10.1016/0038-1098(78)90354-X, http://www.sciencedirect.com/science/article/pii/003810987890354X

Karlsson K, Aryasetiawan F (2000) A many-body approach to spin-wave excitations in itinerant magnetic systems. J Phys Condens Matter 12(34):7617. http://stacks.iop.org/0953-8984/12/i=34/a=308

Khitun A, Wang KL (2005) Nano scale computational architectures with spin wave bus. Superlattice Microst 38(3):184–200. https://doi.org/10.1016/j.spmi.2005.07.001, http://www.sciencedirect.com/science/article/pii/S0749603605000716

Kisker E, Schröder K, Gudat W, Campagna M (1985) Spin-polarized angle-resolved photoemission study of the electronic structure of Fe(100) as a function of temperature. Phys Rev B 31:329–339. https://doi.org/10.1103/PhysRevB.31.329

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Kotani T, van Schilfgaarde M (2008) Spin wave dispersion based on the quasiparticle self-consistent GW method: NiO, MnO and $\alpha$-MnAs. J Phys Condens Matter 20(29):295214. http://stacks.iop.org/0953-8984/20/i=29/a=295214

Kübler J (2009) Theory of itinerant electron magnetism. International series of monographs on physics. OUP, Oxford. https://books.google.de/books?id=ZbM0gHCcmaQC

Loong CK, Carpenter JM, Lynn JW, Robinson RA, Mook HA (1984) Neutron scattering study of the magnetic excitations in ferromagnetic iron at high energy transfers. J Appl Phys 55(6):1895–1897. https://doi.org/10.1063/1.333511

Lounis S, Costa AT, Muniz RB, Mills DL (2010) Dynamical magnetic excitations of nanostructures from first principles. Phys Rev Lett 105:187205. https://doi.org/10.1103/PhysRevLett.105.187205

Lounis S, Costa AT, Muniz RB, Mills DL (2011) Theory of local dynamical magnetic susceptibilities from the Korringa-Kohn-Rostoker Green function method. Phys Rev B 83:035109. https://doi.org/10.1103/PhysRevB.83.035109

Lowde RD, Moon RM, Pagonis B, Perry CH, Sokoloff JB, Vaughan-Watkins RS, Wiltshire MCK, Crangle J (1983) A polarised-neutron scattering demonstration of deviations from stoner-theory behaviour in nickel. J Phys F Metal Phys 13(2):249. http://stacks.iop.org/0305-4608/13/i=2/a=004

Lynn JW (1975) Temperature dependence of the magnetic excitations in iron. Phys Rev B 11:2624–2637. https://doi.org/10.1103/PhysRevB.11.2624

Mahan GD (2000) Many particle physics, 3rd edn. Plenum, New York

Marzari N, Vanderbilt D (1997) Maximally localized generalized Wannier functions for composite energy bands. Phys Rev B 56:12847–12865. https://doi.org/10.1103/PhysRevB.56.12847

Mermin ND, Wagner H (1966) Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. Phys Rev Lett 17:1133–1136. https://doi.org/10.1103/PhysRevLett.17.1133

Minkiewicz VJ, Collins MF, Nathans R, Shirane G (1969) Critical and spin-wave fluctuations in nickel by neutron scattering. Phys Rev 182:624–631. https://doi.org/10.1103/PhysRev.182.624

Monkhorst HJ, Pack JD (1976) Special points for Brillouin-zone integrations. Phys Rev B 13:5188–5192. https://doi.org/10.1103/PhysRevB.13.5188

Mook HA, Nicklow RM (1973) Neutron scattering investigation of the magnetic excitations in iron. Phys Rev B 7:336–342. https://doi.org/10.1103/PhysRevB.7.336

Mook HA, Paul DM (1985) Neutron-scattering measurement of the spin-wave spectra for nickel. Phys Rev Lett 54:227–229. https://doi.org/10.1103/PhysRevLett.54.227

Mook HA, Tocchetti D (1979) Neutron-scattering measurements of the generalized susceptibility $\chi(q, e)$ for Ni. Phys Rev Lett 43:2029–2032. https://doi.org/10.1103/PhysRevLett.43.2029

Moriya T (1985) Spin fluctuations in itinerant electron magnetism. Springer series in solid state sciences, vol 56. Springer, Berlin/Heidelberg

Moruzzi VL, Marcus PM, Schwarz K, Mohn P (1986) Ferromagnetic phases of bcc and fcc Fe, Co, and Ni. Phys Rev B 34:1784–1791. https://doi.org/10.1103/PhysRevB.34.1784

Müller MCTD (2016) Spin-wave excitations and electron-magnon scattering in elementary ferromagnets from ab initio many-body perturbation theory. Ph.D. thesis, RWTH Aachen

Müller MCTD, Friedrich C, Blügel S (2016) Acoustic magnons in the long-wavelength limit: investigating the goldstone violation in many-body perturbation theory. Phys Rev B 94:064433. https://doi.org/10.1103/PhysRevB.94.064433

Müller MCTD, Blügel S, Friedrich C (2019) Electron-magnon scattering in elementary ferromagnets from first principles: lifetime broadening and kinks. Phys Rev B 100:045130

Paul DM, Mitchell PW, Mook HA, Steigenberger U (1988) Observation of itinerant-electron effects on the magnetic excitations of iron. Phys Rev B 38:580–582. https://doi.org/10.1103/PhysRevB.38.580

Rath J, Freeman AJ (1975) Generalized magnetic susceptibilities in metals: application of the analytic tetrahedron linear energy method to SC. Phys Rev B 11:2109–2117. https://doi.org/10.1103/PhysRevB.11.2109

Raue R, Hopster H, Clauberg R (1984) Spin-polarized photoemission study on the temperature dependence of the exchange splitting of Ni. Zeitschrift für Physik B Condens Matter 54(2):121–128. https://doi.org/10.1007/BF01388063

Rosengaard NM, Johansson B (1997) Finite-temperature study of itinerant ferromagnetism in Fe, Co, and Ni. Phys Rev B 55:14975–14986. https://doi.org/10.1103/PhysRevB.55.14975

Rousseau B, Eiguren A, Bergara A (2012) Efficient computation of magnon dispersions within time-dependent density functional theory using maximally localized Wannier functions. Phys Rev B 85:054305. https://doi.org/10.1103/PhysRevB.85.054305

Runge E, Gross EKU (1984) Density-functional theory for time-dependent systems. Phys Rev Lett 52:997–1000. https://doi.org/10.1103/PhysRevLett.52.997

Sakisaka Y, Rhodin T, Mueller D (1985) Angle-resolved photoemission from Fe(110): determination of E(k). Solid State Commun 53(9):793–799. https://doi.org/10.1016/0038-1098(85)90221-2, http://www.sciencedirect.com/science/article/pii/0038109885902212

Santoni A, Himpsel FJ (1991) Unoccupied energy bands, exchange splitting, and self-energy of iron. Phys Rev B 43:1305–1312. https://doi.org/10.1103/PhysRevB.43.1305

Şaşıoğlu E, Schindlmayr A, Friedrich C, Freimuth F, Blügel S (2010) Wannier-function approach to spin excitations in solids. Phys Rev B 81:054434. https://doi.org/10.1103/PhysRevB.81.054434

Şaşıoğlu E, Friedrich C, Blügel S (2013) Strong magnon softening in tetragonal FeCo compounds. Phys Rev B 87:020410. https://doi.org/10.1103/PhysRevB.87.020410

Savrasov SY (1998) Linear response calculations of spin fluctuations. Phys Rev Lett 81:2570–2573. https://doi.org/10.1103/PhysRevLett.81.2570

Scalapino D (1995) The case for $d_{x^2-y^2}$ pairing in the cuprate superconductors. Phys Rep 250(6):329–365. https://doi.org/10.1016/0370-1573(94)00086-I, http://www.sciencedirect.com/science/article/pii/037015739400086I

Schäfer J, Schrupp D, Rotenberg E, Rossnagel K, Koh H, Blaha P, Claessen R (2004) Electronic quasiparticle renormalization on the spin wave energy scale. Phys Rev Lett 92:097205. https://doi.org/10.1103/PhysRevLett.92.097205

Souza I, Marzari N, Vanderbilt D (2001) Maximally localized Wannier functions for entangled energy bands. Phys Rev B 65:035109. https://doi.org/10.1103/PhysRevB.65.035109

Stearns MB (1986) 3D, 4D and 5D elements, alloys and compounds. In: Wijn H, Landolt H, Börnstein R (eds) Magnetic properties in metals. New series, vol III. Springer, Berlin. https://doi.org/10.1007/b29710, http://materials.springer.com/bp/docs/978-3-540-39667-3?utm_campaign=bookshelf-experiment&utm_medium=xls&utm_source=staticpage

Strinati G (1988) Application of the Green's functions method to the study of the optical properties of semiconductors. La Rivista del Nuovo Cimento 11(12):1–86. https://doi.org/10.1007/BF02725962

Tang H, Plihal M, Mills D (1998) Theory of the spin dynamics of bulk Fe and ultrathin Fe(100) films. J Magn Magn Mater 187(1):23–46. https://doi.org/10.1016/S0304-8853(98)00088-2, http://www.sciencedirect.com/science/article/pii/S0304885398000882

Turner AM, Donoho AW, Erskine JL (1984) Experimental bulk electronic properties of ferromagnetic iron. Phys Rev B 29:2986–3000. https://doi.org/10.1103/PhysRevB.29.2986

# Non-equilibrium Green's Function Methods for Spin Transport and Dynamics

Ivan Rungger, Andrea Droghetti, and Maria Stamenova

## Contents

### Abstract

The modeling of spintronic devices is a theoretical challenge, since one has to describe accurately both the electronic structure of the constituent materials and their charge- and spin-transport properties. In this chapter we present the state-of-the-art quantum transport theory appropriate for this task. The theory is based on the so-called non-equilibrium Green's function formalism, which is

I. Rungger (✉)
National Physical Laboratory, Teddington, UK
e-mail: ivan.rungger@npl.co.uk

A. Droghetti
Nano-Bio Spectroscopy Group and European Theoretical Spectroscopy Facility (ETSF),
Materials Physics Center, University of the Basque Country, San Sebastian, Spain
e-mail: andrea.droghetti@ehu.eus

M. Stamenova
School of Physics, AMBER and CRANN Institute, Trinity College, Dublin, Ireland
e-mail: tsonevam@tcd.ie

combined with density functional theory in order to provide a first principles description of materials properties. This allows for the evaluation of the steady-state charge and spin current through a quantum system at a finite applied bias voltage between the electrodes. It also describes the spin-transfer torque that flowing spins exert on localized magnetic moments, which is able to switch the magnetization of a magnetic system. In this chapter the detailed discussion about the principal methodological aspects is accompanied by the review of a number of technologically relevant applications.

# 1 Introduction

In the last decades, the continuous emergence of novel functionalities and related technologies has been a major boost for the advancement of spintronics. The current applications in hard disk drives and magnetic random access memories rely mostly on thin film heterostructures and exploit the magnetoresistance effect for the readout of the magnetic state and the spin-transfer torque for switching it. The key components that determine the functionality are typically only a few nanometers thick, so that the electron transport properties do not obey standard Ohm-Kirchhoff's laws, and a quantum mechanical description is ultimately required. Furthermore, since the device dimensions are often smaller than the electrons' mean free path and phase relaxation length, the coherent, phase-conserving carrier propagation governs the charge and spin transport. As a result, the conductance is determined by the transmission probability across a quantum mechanical scattering region as described in the Landauer-Büttiker framework (Landauer 1957; Büttiker et al. 1985). This requires the calculation of the wave function of the electrons, although the most convenient way to perform simulations of realistic device setups is based on the non-equilibrium Green's function (NEGF) formalism.

In this chapter we describe in detail the NEGF formalism (Stefanucci and van Leeuwen 2013; Datta 1995, 2005; Ghosh 2017), focusing on its application in spintronics. To begin with, we present the general properties of Green's functions for systems at equilibrium. We then discuss open systems and the implementation of NEGF in combination with density functional theory (DFT). This allows to present a number of representative results for first principles NEGF+DFT calculations of conductance, magnetoresistance, and spin-filter properties of nanoscale systems. We then discuss the inclusion of many-electron effects in the formalism and its application to calculate spin-transfer torques and dynamics of nanoscale devices.

# 2 Green's Functions at Equilibrium

A closed quantum system with noninteracting particles is described by the real-space one-particle Hamiltonian operator, $\hat{H}^\sigma(\mathbf{r})$, where the superscript $\sigma = \pm 1/2$ indicates the spin index, and the vector $\mathbf{r}$ is the real space position. Here we assume that the spins are collinear and that $\hat{H}^\sigma(\mathbf{r})$ does not include any spin-mixing term,

which couples up ($\sigma = +1/2$, also denoted by $\uparrow$) and down ($\sigma = -1/2$, also denoted by $\downarrow$) spins. Non-collinear spin systems, necessary also to include the effects of spin-orbit interactions, will be introduced in Sect. 6. The solution of the time-independent Schrödinger equation

$$\hat{H}^\sigma(\mathbf{r})\,\Psi_n^\sigma(\mathbf{r}) = E_n^\sigma\,\Psi_n^\sigma(\mathbf{r}) \tag{1}$$

yields the set of wave functions, $\{\Psi_n^\sigma(\mathbf{r})\}$, and eigenvalues $\{E_n^\sigma\}$ of the system, with the integer $n$ indexing the different eigenstates. Note that this single-particle Hamiltonian can also include electron-electron interactions at a mean-field level, as is the case in semi-local DFT calculations, where the Kohn-Sham (KS) Hamiltonian is used.

The electron density, $\rho(\mathbf{r})$, is obtained as $\rho(\mathbf{r}) = \sum_{n,\sigma} f_n^\sigma |\Psi_n^\sigma(\mathbf{r})|^2$. The sum runs over all eigenstates of $\hat{H}^\sigma$, whose occupation probabilities, $f_n^\sigma = f(E_n^\sigma)$, are given by the Fermi-Dirac distribution:

$$f(E) = \frac{1}{1 + e^{\frac{E - E_{\rm F}}{k_{\rm B}\theta}}}. \tag{2}$$

Here $k_{\rm B}$ is the Boltzmann constant, $\theta$ is the electronic temperature, and $E_{\rm F}$ is the Fermi energy.

In standard quantum transport calculations, $\Psi^\sigma(\mathbf{r})$ is expanded over a local orbitals basis set, $\{\Phi_\beta(\mathbf{r})\}$, as $\Psi^\sigma(\mathbf{r}) = \sum_\beta \psi_\beta^\sigma \Phi_\beta(\mathbf{r})$, where the integer $\beta$ indexes all the orbitals in the basis set, for a total of $N_\beta$ basis orbitals. Typically atomic-like orbitals are used as basis orbitals, where each integer $\beta$ is a collective index for the atom index, the principal quantum number, the angular momentum quantum number, and the magnetic quantum number. In what follows we further assume that each $\Phi_\beta^\sigma(\mathbf{r})$ is non-zero only over a confined region of space. We define the corresponding Hamiltonian matrix, $H^\sigma$, as $H_{\alpha\beta}^\sigma = \langle\Phi_\alpha|\hat{H}^\sigma|\Phi_\beta\rangle = \int \Phi_\alpha^*(\mathbf{r})\hat{H}^\sigma(\mathbf{r})\Phi_\beta(\mathbf{r})d\mathbf{r}$ and the overlap matrix, $\Omega$, as $\Omega_{\alpha\beta} = \langle\Phi_\alpha|\Phi_\beta\rangle = \int \Phi_\alpha^*(\mathbf{r})\Phi_\beta(\mathbf{r})d\mathbf{r}$. With these definitions we can rewrite the Schrödinger equation as $\sum_\beta H_{\alpha\beta}^\sigma \psi_{\beta,n}^\sigma = E_n^\sigma \sum_\beta \Omega_{\alpha\beta} \psi_{\beta,n}^\sigma$, or in the shorter matrix notation $H^\sigma \psi_n^\sigma = E_n^\sigma \Omega \psi_n^\sigma$. Here $\psi_n^\sigma$ is a vector of dimension $N_\beta$, normalized as $\psi_n^{\sigma\dagger}\Omega\psi_n^\sigma = 1$, and $H^\sigma$ and $\Omega$ are Hermitian matrices of dimension $N_\beta \times N_\beta$. The electron density can be expressed as $\rho(\mathbf{r}) = \sum_{\alpha\beta,\sigma} \rho_{\alpha\beta}^\sigma \Phi_\beta^*(\mathbf{r})\Phi_\alpha(\mathbf{r})$, where we have introduced the density matrix, defined as $\rho_{\alpha\beta}^\sigma = \sum_n f_n^\sigma \psi_{\alpha,n}^\sigma \psi_{\beta,n}^{\sigma\,*}$. It can be written equivalently in the shorter matrix notation as $\rho^\sigma = \sum_n f_n^\sigma \psi_n^\sigma \psi_n^{\sigma\dagger}$. The total number of electrons, $N_e = N_e^\uparrow + N_e^\downarrow$, is then obtained from $N_e^\sigma = \int \rho^\sigma(\mathbf{r})d\mathbf{r} = \sum_{\alpha\beta} \rho_{\alpha\beta}^\sigma \Omega_{\beta\alpha} = \text{Tr}[\rho^\sigma \Omega] = \sum_n f_n^\sigma$. For a given $N_e$ this relation implicitly determines $E_{\rm F}$.

The energy-dependent single-particle retarded Green's function (GF), $G^\sigma(E)$, is defined as

$$G^\sigma(E) = [(E + i\delta)\Omega - H^\sigma]^{-1}, \tag{3}$$

where $\delta \rightarrow 0^+$. If the eigenstates of $H^\sigma$ are known, then it can be written in its spectral form as $G^\sigma(E) = \sum_{n=1}^{N_\beta} \frac{1}{E+i\delta-E_n^\sigma} \psi_n^\sigma \psi_n^{\sigma\,\dagger}$. We note that if $H^\sigma$ and $\Omega$ are both real, then also the $\psi_n^\sigma$ can be made real, so that $G^\sigma(E)$ is symmetric. We now define the spectral function, $A^\sigma(E)$, as

$$A^\sigma(E) = i\left[G^\sigma(E) - G^{\sigma\,\dagger}(E)\right]. \tag{4}$$

By expanding $G^\sigma(E)$ in its spectral form, $A^\sigma(E)$ can be rewritten as

$$A^\sigma(E) = \lim_{\delta\rightarrow 0^+} 2\sum_{n=1}^{N_\beta} \frac{\delta}{(E-E_n^\sigma)^2+\delta^2} \psi_n^\sigma \psi_n^{\sigma\,\dagger} = 2\pi \sum_{n=1}^{N_\beta} \delta(E-E_n^\sigma)\,\psi_n^\sigma \psi_n^{\sigma\,\dagger}, \tag{5}$$

which shows that $A^\sigma(E)$ has a peak at each eigenstate of $H^\sigma$ and therefore corresponds to a generalized density of states (DOS). With above relation the spin-dependent density matrix, $\rho^\sigma = \sum_n f_n^\sigma \psi_n^\sigma \psi_n^{\sigma\,\dagger}$, can be written as

$$\rho^\sigma = \frac{1}{2\pi} \int dE\, f(E)\, A^\sigma(E), \tag{6}$$

where the energy integral goes from $-\infty$ to $+\infty$. If one defines the lesser Green's function, $G^{\sigma\,<}(E)$, for systems in equilibrium as

$$G^{\sigma\,<}(E) = if(E)\,A^\sigma(E), \tag{7}$$

then $\rho^\sigma$ can be rewritten as

$$\rho^\sigma = \frac{1}{2\pi i} \int dE\, G^{\sigma\,<}(E). \tag{8}$$

In Green's function-based numerical calculations, one therefore needs to calculate $G^{\sigma\,<}(E)$ and integrate this quantity over energies in order to then obtain the electron density.

In this chapter the Green's function method is combined with the KS equations of DFT, unless explicitly stated otherwise. Therefore, following a standard practice, we identify the single-particle Hamiltonian $\hat{H}^\sigma$ with the KS Hamiltonian, $\hat{H}_{KS}^\sigma$, and the corresponding eigenstates with the KS states. $\hat{H}_{KS}^\sigma$ includes a single-particle kinetic energy term, external classical potentials (such as the electron-nuclei interaction), the Hartree term describing the classical Coulomb interaction between electrons, and the exchange-correlation potential, which accounts for all remaining quantum many-body contributions. Spin-transport calculations are generally performed by using a semi-local approximation to the exchange-correlation density functional, such as the local spin density approximation (LSDA) (von Barth and Hedin 1972). DFT is a variational many-body theory that enables the calculation of the ground

state observables of a system, in particular the total energy, from the charge density, but the KS eigenvalues do not correspond to the true excitation energies. In exact DFT the only exception is the eigenenergy of the highest occupied state, which is equal to minus the ionization energy (Perdew and Levy 1997), although within LSDA-DFT this relation is only approximately valid. Nevertheless, for many systems the KS states provide a good approximate description of the electronic structure. Typically the physical nature of the studied system, e.g., whether it is metallic or insulating, as well as the orbital character of the valence and conduction states, is correctly predicted, although the calculated energy gaps or bandwidths deviate from experiments.

The quantitative shortcomings of KS-LSDA are often related to the self-interaction error (Perdew and Zunger 1981), which tends to favor charge delocalization and shifts localized states to higher energies. A number of exchange-correlation functionals to correct this error have been proposed. In quantum transport calculations, an efficient correction is given by the so-called atomic self-interaction correction (ASIC) (Pemmaraju et al. 2007) and by the LSDA+U method (Himmetoglu et al. 2014). ASIC corrects the self-interaction error in an approximate way for all occupied states, while it does not affect empty states, which have no self-interaction. In LSDA+U a Hubbard-like $U$ potential is applied to a chosen subset of either atomic orbitals, typically the $3d$ and $4f$ states, or molecular orbitals, which are expected to be localized due to strong correlation effects. This potential discourages fractional occupations and acts differently on occupied and unoccupied orbitals, leading to a bandgap of the order of $U$. The $U$ parameter, which describes the screened rather than the bare Coulomb interaction, is typically fitted so to achieve a satisfactory agreement with the experimental results, although different methods to calculate it from first principles have been proposed (for a review see Himmetoglu et al. 2014). In case of molecules or other spacers between metal electrodes, the effect of screening can be accounted for by using an image-charge model (Souza et al. 2014; Pertsova et al. 2015), while for $d$ or $f$ orbitals in solids several schemes based either on constrained DFT (Anisimov and Gunnarsson 1991), constrained random phase approximation (Miyake et al. 2009), and linear response (Himmetoglu et al. 2014) have been proposed.

## 3 Non-equilibrium Green's Function for Quantum Transport

The GF method described in the previous section can be naturally extended to describe out-of-equilibrium open systems at steady state. The typical system that we consider represents two-terminal devices at an applied bias voltage. It is subdivided in three parts: a central scattering region, also called extended molecule (EM), and left and right electrodes, from which electrons flow in and out (Fig. 1). We use the convention that the $z$-axis is the transport direction, and we assume that each electrode is semi-infinite and periodic away from the EM along $z$. The EM and the left (right) electrodes are described by the Hamiltonian matrices $H_{\text{EM}}^{\sigma}$ and
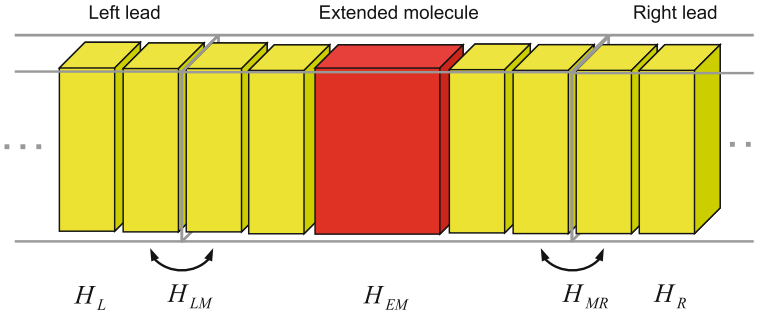
**Fig. 1** Schematic representation of the two-terminal setup used in typical quantum transport calculations. A central scattering region, also denoted as extended molecule (EM), is joined to the left and right leads, also denoted as electrodes. In practical quantum transport calculations, the EM typically includes at least one unit cell of the semi-infinite periodic electrodes at each of its ends, although in general this condition is not strictly necessary

$H_{L(R)}^\sigma$, respectively. The coupling between the left (right) electrode and the EM is accounted for by $H_{LM(MR)}^\sigma$, while there is no direct coupling between the two electrodes (Fig. 1). Due to its infinitely large size, each electrode is in local thermal equilibrium at its own chemical potential, $\mu_L$ for the left electrode and $\mu_R$ for the right electrode. The applied bias voltage is then defined as $V_b = \mu_L - \mu_R$. Note that the units of $V_b$ are in eV, and the corresponding voltage in volt, $V = V_b/e$, is obtained by dividing $V_b$ by the electron charge, $e$ ($e = -|e|$). The values of $\mu_{L(R)}$ are defined up to an arbitrary constant shift of the whole energy spectrum, and here we implicitly define this constant by setting $\mu_L = E_F + V_b/2$ and $\mu_R = E_F - V_b/2$.

A crucial advantage of solving the transport problem by using the Green's function method is that the EM can be separated out from the semi-infinite leads. Their effect on the EM is then captured via the energy-dependent retarded self-energies of the leads, denoted as $\Sigma_L^\sigma(E, V_b)$ ($\Sigma_R^\sigma(E, V_b)$) for the left (right) electrode. In analogy to Eq. (3), this allows to write the retarded GF of the EM at an applied bias voltage as

$$G_{EM}^\sigma(E, V_b) = \left[ (E + i\delta)\Omega_{EM} - H_{EM}^\sigma - \Sigma_L^\sigma(E, V_b) - \Sigma_R^\sigma(E, V_b) \right]^{-1}, \quad (9)$$

where $\Omega_{EM}$ is the overlap matrix of the EM. The detailed procedure to evaluate $\Sigma_L^\sigma(E, V_b) = \Sigma_L^\sigma(E - V_b/2, 0)$ and $\Sigma_R^\sigma(E, V_b) = \Sigma_R^\sigma(E + V_b/2, 0)$ goes beyond the scope of this chapter; the established methods for their calculation are based on either a semi-analytical expression (Rungger and Sanvito 2008) or on iterative procedures (López Sancho et al. 1984). The leads' self-energies, $H_{EM}$, $\Omega_{EM}$, and $G_{EM}^\sigma$, are all matrices of dimension $N_{EM} \times N_{EM}$, where $N_{EM}$ is the number of basis orbitals in the EM. Importantly, the self-energies are not Hermitian, and their antihermitian parts describe the flow of electrons between EM and the leads. To quantify these antihermitian components, one introduces the matrices

$\Gamma^\sigma_{L(R)} = i \left( \Sigma^\sigma_{L(R)} - \Sigma^\sigma_{L(R)}{}^\dagger \right)$, which are positive semi-definite and represent the energy-dependent strength of the electronic coupling of the EM to each of the electrodes. This can be seen by writing the spectral function from Eq. (4) as $A^\sigma = A^\sigma_L + A^\sigma_R$, with $A^\sigma_{L(R)} = G^\sigma_{EM} \Gamma^\sigma_{L(R)} G^\sigma_{EM}{}^\dagger$, which can be verified using Eq. (9), and assuming that there are no fully localized states in the EM, so that one can take $\delta$ to 0. The states of the system are therefore the sum of those coupled to the left electrode, which determine $A^\sigma_L$, and those coupled to the right electrode, which determine $A^\sigma_R$. In the out-of-equilibrium situation, the states coupled to the left (right) electrode are therefore at the chemical potential $\mu_L$ ($\mu_R$). This allows us to generalize the lesser GF from Eq. (7) to the non-equilibrium situation as $G^{\sigma <} = i f_L A^\sigma_L + i f_R A^\sigma_R$, where $f_{L(R)}(E, V_b) = f(E - \mu_{L(R)})$. If one defines the lesser lead self-energies, $\Sigma^\sigma_{L(R)}{}^<(E, V_b)$, as

$$\Sigma^\sigma_{L(R)}{}^<(E, V_b) = i f_{L(R)}(E, V_b) \Gamma^\sigma_{L(R)}(E, V_b), \tag{10}$$

this lesser GF is equivalently given by

$$G^\sigma{}^<_{EM}(E, V_b) = G^\sigma_{EM}(E, V_b) \left[ \Sigma^\sigma_L{}^<(E, V_b) + \Sigma^\sigma_R{}^<(E, V_b) \right] G^\sigma_{EM}{}^\dagger(E, V_b). \tag{11}$$

The density matrix of the EM, $\rho^\sigma_{EM}$, can then be evaluated at each bias by using Eq. (8):

$$\rho^\sigma_{EM}(V_b) = \frac{1}{2\pi i} \int dE \, G^\sigma{}^<_{EM}(E, V_b), \tag{12}$$

where we have explicitly introduced the energy and voltage dependence of all quantities. Furthermore, we also define the energy density matrix, $W^\sigma$, which is required for the calculation of forces and spin currents (see Sect. 6) and which is given by

$$W^\sigma(V_b) = \frac{1}{2\pi i} \int dE \, E \, G^\sigma{}^<(E, V_b). \tag{13}$$

We note that for computational efficiency the integral in the last two equations is usually split in a part running over real energies and a part evaluated over complex energies (Rocha et al. 2006).

When we use the DFT KS Hamiltonian for $H^\sigma_{EM}$, this becomes a functional of the density, i.e., $H^\sigma_{EM}(V_b) = H^\sigma_{EM}[\rho^\uparrow(V_b), \rho^\downarrow(V_b)]$. In this case $\rho^\sigma(V_b)$ and $H^\sigma_{EM}(V_b)$ need to be solved self-consistently, in analogy to standard ground state DFT calculations. This is typically achieved in an iterative procedure, where at each iteration for a given input $\rho^\sigma(V_b)$, one computes $H^\sigma_{EM}(V_b)$, which is then used to evaluate $G^\sigma_{EM}$ and with it to obtain a new output $\rho^\sigma(V_b)$. Self-consistency is achieved once the difference between the input and output density matrices is below

a given tolerance. Note that at finite $V_b$ we do not have a variational principle for the energy, so that at self-consistency instead of the ground state we obtain a steady state, where the inflow of electrons equals the outflow.

Once $\rho^\sigma(V_b)$ is calculated self-consistently, one can obtain the current across the EM, $I = I^\uparrow + I^\downarrow$, as

$$I^\sigma = \frac{G_0}{2e} \int T^\sigma(E, V_b) \left[ f_L(E, V_b) - f_R(E, V_b) \right] dE, \tag{14}$$

where $G_0 = 2e^2/h$ is the quantum of conductance, $h$ is the Planck constant, and $T^\sigma(E, V_b)$ is the spin-dependent transmission coefficient. This is given by

$$T^\sigma(E, V_b) = \mathrm{Tr} \left[ \Gamma_L^\sigma(E, V_b) G_{EM}^{\sigma}{}^\dagger(E, V_b) \Gamma_R^\sigma(E, V_b) G_{EM}^\sigma(E, V_b) \right], \tag{15}$$

where "Tr" denotes the matrix trace operation. The integrand in Eq. (14) is large only in an energy range of about $\pm V_b/2$ around $E_F$, which is called the bias window. The current is therefore mainly determined by the transmission integrated across this bias window.

If periodic boundary conditions are used in the $x - y$ plane perpendicular to the transport direction $z$, then $H_{EM}^\sigma$, $\Omega_{EM}$, $G_{EM}^\sigma$, $G_{EM}^\sigma{}^<$, $\Sigma_{L/R}^\sigma$, and $\Gamma_{L/R}^\sigma$ acquire a $\mathbf{k} = (k_x, k_y)$ index, so that we now have $H_{\mathbf{k},EM}^\sigma$, $\Omega_{\mathbf{k},EM}^\sigma$, $G_{\mathbf{k},EM}^\sigma(E, V_b)$, $G_{\mathbf{k},EM}^\sigma{}^<(E, V_b)$, $\Sigma_{\mathbf{k},L/R}^\sigma(E, V_b)$, and $\Gamma_{\mathbf{k},L/R}^\sigma(E, V_b)$. The equations outlined above are then valid for each $\mathbf{k}$-point, and the total density matrix and transmission are given by an integral over the Brillouin zone (BZ) perpendicular to the transport direction as

$$\rho_{EM}^\sigma(E, V_b) = \frac{1}{\Omega_{BZ}} \int_{\Omega_{BZ}} \rho_{\mathbf{k},EM}^\sigma(E, V_b) dk_x dk_y \tag{16}$$

and

$$T^\sigma(E, V_b) = \frac{1}{\Omega_{BZ}} \int_{\Omega_{BZ}} T_{\mathbf{k}}^\sigma(E, V_b) dk_x dk_y \tag{17}$$

where $\Omega_{BZ}$ is the area of the BZ.

The method described in this section is referred to as the NEGF formalism. When one uses the DFT KS Hamiltonian in the calculation, it is usually referred to as the NEGF+DFT method. In the following section, we will show how NEGF+DFT allows for the description of spin-dependent transport properties in nanoscale devices from first principles. All results are obtained by using the *Smeagol* quantum transport code (Rocha et al. 2006; Rungger and Sanvito 2008), which gets as input the DFT Hamiltonian from the *Siesta* DFT code (Soler et al. 2002).

# 4       Magnetoresistance and Spin Filtering

Thin film heterostructures combining magnetic and nonmagnetic materials are the most widely used spintronic devices to date. The typical setup, called spin valve, consists of two magnetic metals separated by one or more nonmagnetic spacer layers, which can be insulating, semiconducting, or metallic (Fig. 2a, b, d). For such systems the key quantity that determines the device performance is the magnetoresistance (MR), which is defined as

$$\text{MR} = \frac{I^{\text{P}} - I^{\text{AP}}}{I^{\text{small}}}, \tag{18}$$

where $I^{\text{P}}$ is the current for parallel (P) alignment of the magnetizations of the electrodes, $I^{\text{AP}}$ is the current for antiparallel (AP) alignment, and $I^{\text{small}}$ is the smaller of these two currents. The MR can be above tens of percent for metallic junctions, as first shown by A. Fert and P. Grünberg with their independent discovery of the giant magnetoresistance (GMR) effect (Baibich et al. 1988; Binasch et al. 1989). For insulating spacers the tunneling MR (TMR) can reach up to hundreds or thousands of percent, as is the case for the now widely used Fe/MgO/Fe(100) and related tunnel junctions. Notably, such huge TMR was first predicted by using NEGF+DFT (Butler et al. 2001) and subsequently confirmed experimentally (Parkin et al. 2004; Yuasa et al. 2004).

The general qualitative behavior of the MR for thin film stacks can be understood by modeling the left ferromagnet/spacer and right spacer/ferromagnet interfaces in
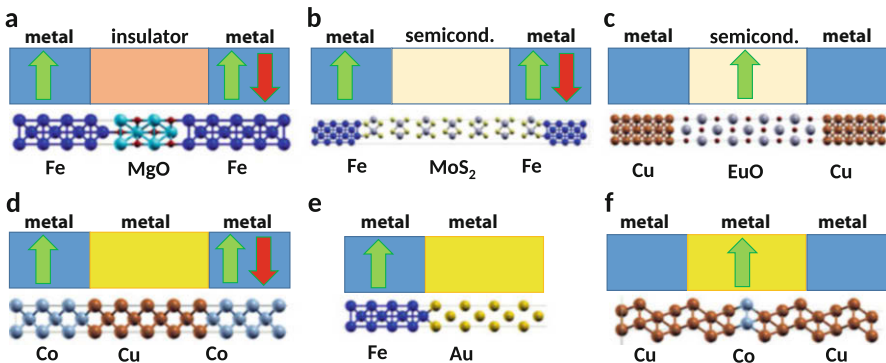


**Fig. 2** Prototypical spintronic device setups involving various combinations of magnetic and nonmagnetic layers, as well as metallic and insulating layers. The arrows indicate that a layer is magnetic, and the presence of a red down arrow indicates systems used for their magnetoresistance, obtained by reversing the magnetization direction of one of the layers. The unit cells shown underneath each general system setup indicate atomic structures of specific examples of its practical realization
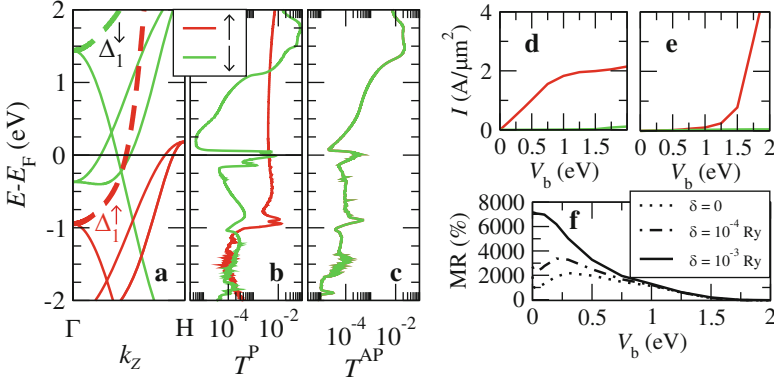
**Fig. 3** (**a**) Fe band structure along the (100) direction, where the thick bands indicate states with $\Delta_1$ symmetry; spin-polarized transmission for P (**b**) and AP (**c**) alignment of the magnetization of the Fe electrodes in a Fe/MgO/Fe(100) junction with four MgO monolayers and resulting spin-dependent current for P (**d**) and AP (**e**) alignment; the resulting magnetoresistance as function of bias voltage is shown in (**f**). (Figure adapted from Rungger et al. 2009)

the devices of Fig. 2a, b, d as two independent scatters in series. If one assumes symmetric interfaces and neglects interference effects across interfaces, then one can indicate the majority (minority) spin transmission of each individual interface as $T^M(E)$ [$T^m(E)$]. Since spin-mixing is neglected, the P transmission can then be approximated as $T^P = (T^M)^2 + (T^m)^2$, while the AP transmission is $T^{AP} = 2T^m T^M$. For systems with $T^M \gg T^m$, indicating spin-filtering interfaces, one therefore has $T^P \gg T^{AP}$, which implies $I^P \gg I^{AP}$ and with it a very large MR. Even though this simple model does not apply to all systems (De Teresa et al. 1999), the general requirement for large MR is indeed that for the P alignment the current is highly spin-polarized.

In Fig. 3b we show a representative spin-polarized transmission curve of a Fe/MgO/Fe(100) junction as a function of the energy for the P alignment of the magnetic electrodes (Rungger et al. 2009). In the energy range from $-1$ to $+1\,\text{eV}$ around $E_F$ ones has $T^\uparrow \gg T^\downarrow$, and therefore this junction satisfies the requirement for large MR . Here the up (down) spin transmission is the majority (minority) transmission, so that $T^\uparrow = T^M$ and $T^\downarrow = T^m$. The AP transmission (Fig. 3c) lies between $T^\uparrow$ and $T^\downarrow$ calculated for the P alignment, in agreement to the qualitative model explained above. The origin of the spin-filtering property of this junction is the fact that MgO allows high transmission only for states with $\Delta_1$ symmetry (e.g., $s$-states, $p_z$-states, or $d_{z^2}$-states), and these states are only found in the up-spin spectrum of the Fe electrodes around $E_F$ (Fig. 3a). The current versus voltage characteristic curve ($I$-$V$) for P alignment therefore shows that $I^\uparrow \gg I^\downarrow$ at a moderate bias (Fig. 3d). In contrast, for the AP alignment, both $I^\uparrow$ and $I^\downarrow$ are very small at the same moderate bias, so that one has $I^P \gg I^{AP}$ (Fig. 3d, e) and a very high TMR (Fig. 3f). Note that $T^\downarrow$ for P alignment has a sharp peak very close to $E_F$ due to the formation of an interface state between Fe and MgO in the
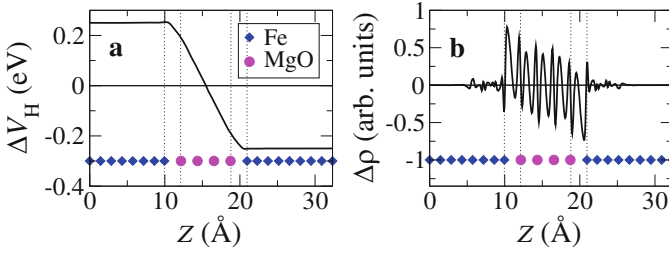
**Fig. 4** Averages over the $(x, y)$ plane of the difference between finite-bias ($V_b = 0.5\,\text{eV}$) and zero-bias potential (**a**) and charge density (**b**) as function of the position along the transport direction $z$ in the Fe/MgO/Fe(100) junction. (Figure reproduced from Rungger et al. 2009)

down spins. With its high DOS, it leads to an increased $T^\downarrow$ in a small energy range around $E_F$, which does however become negligible compared to $T^\uparrow$ as the MgO thickness increases beyond the four monolayers considered here. Furthermore, the contribution of the interface states is drastically reduced when considering a small amount of disorder at the interface, which to a first approximation is achieved by applying a finite $\delta$ in the calculation of $G_{EM}$ (Eq. (9)). This leads to a further suppression of the low bias $I^{AP}$ and with it an increase of the TMR. For larger $V_b$ the TMR decreases progressively, since at about 1 eV $T^\downarrow$ becomes larger than $T^\uparrow$ (Fig. 3b). We note that the DFT bandgap of MgO is smaller than the experimental one (4.7 eV for the used LSDA vs. 7.8 eV in experiment). We can then correct the MgO DFT gap by using ASIC to achieve a quite good agreement with the experimental results. While ASIC changes the detailed quantitative values of the current and TMR, it leaves the qualitative picture obtained from LSDA unaffected. We also note that in these calculations the charge density and potential have been computed self-consistently at finite bias, and the typically resulting potential drop and charge polarization are shown in Fig. 4. One can see that the potential drop is approximately linear between the two metal electrodes, so that to a first approximation, one can also simulate the finite-bias potential non-self-consistently by rigidly shifting it inside the metal electrodes and by applying a linear potential ramp between them. The resulting transmission and current are similar to the ones obtained in the self-consistent calculation.

The key requirements in order to obtain high TMR are the good crystallinity of the interfaces and the large bandgap of the spacer. Both these requirements are met in Fe/MgO/Fe(100) junctions, which exhibit good epitaxial growth. In contrast, the smaller band gap of semiconducting spacers usually results in metal/semiconductor/metal junctions having a lower MR. For instance, in Fe/MoS$_2$/Fe junctions (Fig. 2b), multilayer MoS$_2$ has a gap of about 1–2 eV, and the epitaxy with Fe is less good. The MR is therefore reduced from thousands to a few hundred percent (Fig. 5a), which is nevertheless still sizeable (Dolui et al. 2014). Due to the smaller bandgap of MoS$_2$, the decay of MR with bias is also faster, and the MR approximately vanishes at a voltage of about 0.4 V (Fig. 5b). For metallic spacers the assumption used in the presented NEGF calculations of a fully coherent
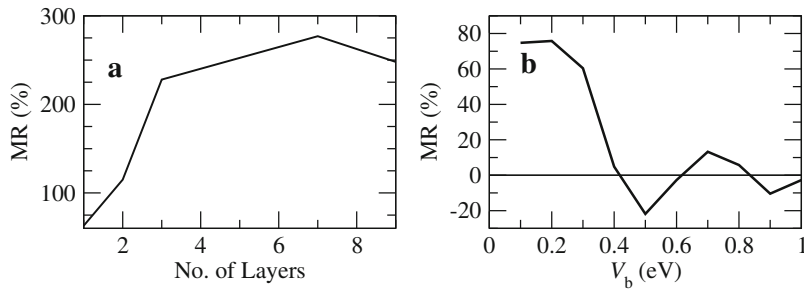
**Fig. 5** (**a**) Magnetoresistance for a Fe/MoS$_2$/Fe(100) junction as function of the number of MoS$_2$ layers; (**b**) bias voltage dependence of the MR for a seven monolayer MoS$_2$ junction. (Figure adapted from Dolui et al. 2014)

current is not valid, since diffusive contributions to the electron transport become important even for thin spacers. The evaluation of the relative contribution of diffusive and coherent (also called ballistic) currents is an active area of theoretical and experimental research. Using ultrashort laser pulses on Fe/Au films (Fig. 2e), it is possible to estimate their relative contributions, and it is found that high-energy ballistic electrons can travel for tens of nanometers and that the NEGF transmission coefficient across the Fe/Au interface plays a determining role in the magnitude of the currents and of their spin polarization (Alekhin et al. 2017).

The example of the Fe/MgO/Fe(100) junction has demonstrated that the key factor to obtain a large MR is that each magnetic electrode in conjunction with the specific spacer acts as spin filter. Such spin filters can also be obtained by placing a ferromagnetic layer, either metallic or semiconducting, between nonmagnetic metallic electrodes (Fig. 2c, f). For common applications the main drawback of all known magnetic semiconductors is that they have rather low Curie temperatures far below room temperature. Despite that, ferromagnetic semiconductors can still be used in devices operating in cryogenic conditions. As typical example of ferromagnetic semiconductor we consider EuO. Here there are only spin-up states at the top of the valence band and at the bottom of the conduction band, which fix the bandgap. In contrast, the down-spin band structure shows a much larger gap. Consequently, in a prototypical Cu/EuO/Cu junction, the total transmission around $E_F$ is mostly determined by $T^\uparrow$ (Fig. 6), while $T^\downarrow$ is much smaller (Jutong et al. 2012). EuO therefore acts as an almost perfect spin filter, where the spin polarization of the current remains large up to high bias. Note that in these calculations the LSDA+U is used for the Eu $4f$ states, since the LSDA bandgap is found to be vanishingly small, in contrast to experimental observations (Pertsova et al. 2015).

Finally, we point out that spintronic devices can be realized also by using single molecules or molecular layers contacted between metal electrodes. For example, in the work by Rocha et al. (2005), a spin valve consisting of a 1,4-[3]-phenyl-dithiolate (tricene-dithiolate) molecule between Ni electrodes is investigated by using the NEGF method. The molecule corresponds to a nonmagnetic spacer,
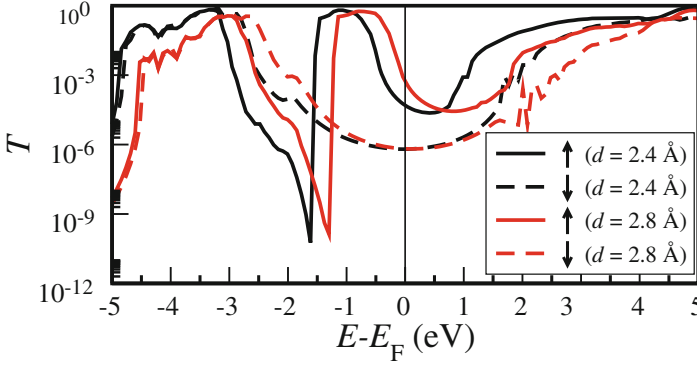
**Fig. 6** Spin-resolved transmission for a Cu/EuO/Cu junction, for two different distances, $d$, between Cu and EuO. The transmission around $E_F$ is dominated by the majority (up) spins, and the precise alignment of the transmission spectrum with respect to $E_F$ depends on the distance between Cu and EuO. (Figure adapted from Jutong et al. 2012)

and the transmission function becomes spin-dependent as a result of the strong electronic coupling between the molecule and the spin-polarized Ni $d$ orbitals. The theoretical MR reaches about 600% for a bias voltage of 0.5 V.

## 5    Electron Correlations Beyond Kohn-Sham DFT

The successful application of DFT+NEGF to the nanodevices presented so far relies on the fact that the LSDA KS states return a reasonably accurate description of the electronic structure. As we have indicated for the Fe/MgO/Fe and Cu/EuO/Cu junctions, mean-field (MF) corrections such as the ASIC and LSDA+U can often bring DFT results in agreement with experiments. However, there are a number of effects, which are due to electron-electron correlations not captured by means of any static mean-field approach. In these cases, the KS Green's function can be used as zeroth-order approximation within a more accurate treatment based on advanced many-body techniques. The extension of the NEGF method to many-body systems is formally accomplished by adding to Eqs. (9) and (11) the retarded and lesser self-energies for the many-body (MB) electron-electron interaction, $\Sigma_{MB}^{\sigma}$ and $\Sigma_{MB}^{\sigma}{}^{<}$. Thus, one defines the MB Green's functions as

$$G_{MB}^{\sigma}(E, V_b) = \big[ (E + i\delta)\Omega_{EM} - H_{EM}^{\sigma} - \Sigma_L^{\sigma}(E, V_b) - \Sigma_R^{\sigma}(E, V_b) \\ - \Sigma_{MB}^{\sigma}(E, V_b) \big]^{-1}, \tag{19}$$

$$G_{MB}^{\sigma}{}^{<}(E, V_b) = G_{MB}^{\sigma}(E, V_b) \big[ \Sigma_L^{\sigma}{}^{<}(E, V_b) + \Sigma_R^{\sigma}{}^{<}(E, V_b) \\ + \Sigma_{MB}^{\sigma}{}^{<}(E, V_b) \big] G_{MB}^{\sigma}{}^{\dagger}(E, V_b). \tag{20}$$

We note that, if $H_{\rm EM}$ is the KS Hamiltonian, then the contribution to the electron correlation already accounted for at the KS level has to be subtracted out of $\Sigma_{\rm MB}$ and $\Sigma_{\rm MB}^<$. This so-called double counting correction is unfortunately unknown in general and has to be approximated (Kotliar et al. 2006). Also other interactions, such as the electron-phonon interaction, can be formally included by introducing into these equations the corresponding self-energies, which just add up to the leads' self-energies. However, we remark that a key assumption to guarantee the additive property of the interaction self-energies with the leads' self-energies is that these interactions are localized in the EM.

The formal introduction of the electron-electron interaction via self-energies in Eqs. (19) and (20) shows that the interaction effectively acts on the system as an additional electrode. Then one can define an effective "coupling matrix" for the interaction, $\Gamma_{\rm MB}^\sigma(E, V_{\rm b}) = i\left(\Sigma_{\rm MB}^\sigma(E, V_{\rm b}) - \Sigma_{\rm MB}^{\sigma \dagger}(E, V_{\rm b})\right)$, and express $\Sigma_{\rm MB}^{\sigma\,<}(E, V_{\rm b})$ in terms of $\Gamma_{\rm MB}^\sigma(E, V_{\rm b})$ through the relation (Droghetti and Rungger 2017; Ness et al. 2010):

$$\Sigma_{\rm MB}^{\sigma\,<}(E, V_{\rm b}) = i\, F_{\rm MB}^\sigma(E, V_{\rm b})\Gamma_{\rm MB}^\sigma(E, V_{\rm b}). \tag{21}$$

This has a similar structure to Eq. (10). However, we note that $F_{\rm MB}^\sigma$ is a $N_{\rm EM} \times N_{\rm EM}$ distribution matrix, which contains information about the interacting out-of-equilibrium in- and outflow of electrons from the EM. The calculation of $F_{\rm MB}^\sigma$ remains a challenge to date and is the topic of ongoing research. Only in equilibrium (i.e., for $V_{\rm b} = 0$) $F_{\rm MB}^\sigma$ becomes the Fermi-Dirac distribution.

The current can be expressed as (Droghetti and Rungger 2017; Ness et al. 2010)

$$I^\sigma = \frac{G_0}{2e}\int\left\{T^\sigma(E, V_{\rm b})\left[f_{\rm L}(E) - f_{\rm R}(E)\right] + {\rm Tr}\Big[[F_{\rm MB}^\sigma(E, V_{\rm b}) - f_{\rm R}(E)]\right.$$
$$\left.\Gamma_{\rm MB}^\sigma(E, V_{\rm b})G_{\rm MB}^{\sigma\,\dagger}(E, V_{\rm b})\Gamma_{\rm R}^\sigma(E, V_{\rm b})G_{\rm MB}^\sigma(E, V_{\rm b})\Big]\right\}dE. \tag{22}$$

$T^\sigma(E, V_{\rm b})$ is the transmission coefficient defined as in Eq. (15), but here it is computed by using the MB retarded Green's function in Eq. (19), i.e.:

$$T^\sigma(E, V_{\rm b}) = {\rm Tr}\left[\Gamma_{\rm L}^\sigma(E, V_{\rm b})G_{\rm MB}^{\sigma\,\dagger}(E, V_{\rm b})\Gamma_{\rm R}^\sigma(E, V_{\rm b})G_{\rm MB}^\sigma(E, V_{\rm b})\right]. \tag{23}$$

The last term on the right-hand side of Eq. (22) accounts for the effect of the "interaction electrode." It is similar to the first one, but $F_{\rm EM}^\sigma(E, V_{\rm b})$ cannot be brought outside the trace, and we cannot define a transmission coefficient related to the flow of electrons from the left lead into the interaction lead and then from this to the right lead. The first term of Eq. (22) accounts for the elastic contributions to the current, while the second term accounts for the inelastic ones, which are often not negligible. $\Gamma_{\rm MB}^\sigma$ is usually non-zero except for a few special cases such as the so-called irregular self-energies. These are instantaneous (local) in time self-energies, such as the Hartree-Fock one. $\Gamma_{\rm MB}^\sigma$ also vanishes for Fermi liquid systems at zero temperature, but only at the Fermi energy. Away from it there is always a non-zero contribution coming from the continuum of electron-hole excitations.

To date, studies including correlation effects for spin-dependent transport beyond NEGF+DFT are scarce and mostly limited to molecular devices and point contacts at zero-bias (see, e.g., Jacob (2015) and references therein) rather than to solid-state systems. In fact, although the discussion above provides the formal equations to solve the problem, in practice the calculation of the MB retarded and lesser self-energies within a treatable approximation is a challenging task. One possible approach is to use the GW approximation of the many-body perturbation theory to obtain a corresponding MB self-energy (Strange et al. 2011). A second approach is represented by the dynamical mean-field theory (DMFT) (Kotliar et al. 2006). In DMFT one starts by selecting a subspace of correlated orbitals in the very same way as in LSDA+U. The effective low-energy Hamiltonian of this subspace is projected out from the whole system and then solved by using DMFT to obtain the corresponding MB self-energy. Finally this is projected back from the correlated subspace to the original system. One of the strengths of DMFT is that it is not a perturbative approach, so that both weak and strong correlations are treated on an equal footing. However, DMFT neglects spatially nonlocal correlations.

As an example of application of NEGF+DFT+DMFT to a solid-state system, we present results for a Cu/monolayer-Co/Cu heterostructure (Chioncel et al. 2015) in Fig. 7. We note that to date no first principles study at finite bias has been reported and the transport properties have been assessed by calculating only the elastic transmission as defined in Eq. (23). DMFT significantly changes the DOS of the Co layer when compared to KS-LSDA. The most striking effect is the narrowing of the Co-3$d$ states and the reduced exchange splitting, and this is reflected in the transmission function. In fact, although its order of magnitude around $E_F$ is approximately unchanged, the spin polarization is significantly reduced in DMFT compared to LSDA results.
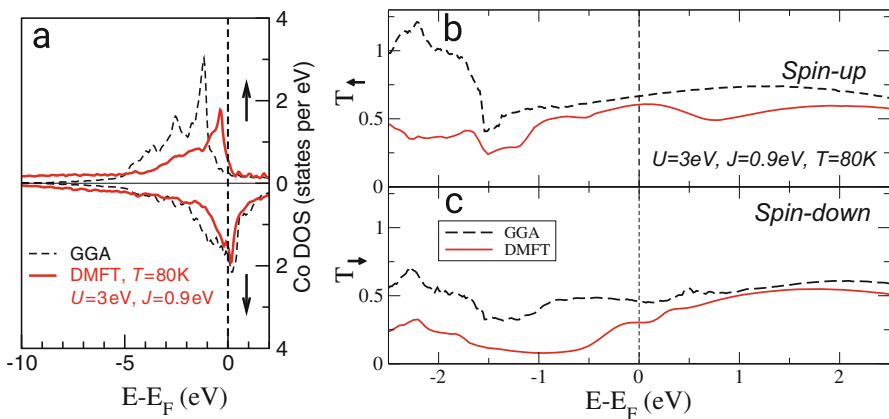


**Fig. 7** (**a**) Density of states of the Co layer between Cu electrodes and (**b**) up- and (**c**) down-spin transmission as function of energy. Black dashed curves (red solid lines) represent results without (with) DMFT corrections. (Figure adapted from Chioncel et al. 2015)

## 6 Finite-Bias Spin-Transfer Torque

If spin-mixing terms are present in the Hamiltonian, $H^\uparrow$ and $H^\downarrow$ are not anymore independent of each other. The Hamiltonian matrix then needs to be rewritten in a general non-collinear form as (Xie et al. 2016)

$$H = H_0 \otimes 1_2 + H_x \otimes \sigma_x + H_y \otimes \sigma_y + H_z \otimes \sigma_z \tag{24}$$

where $1_2$ is the $2 \times 2$ identity matrix, $\sigma_{x/y/z}$ are the three $2 \times 2$ Pauli spin matrices, and $H_{0/x/y/z}$ are matrices, which are of general dimension $N_\beta \times N_\beta$ (for an open system $N_\beta \to \infty$). For each pair of orbitals $\alpha$ and $\beta$ we have a $2 \times 2$ spin block matrix $H_{\alpha\beta} = H_{0,\alpha\beta} \, 1_2 + \mathbf{H}_{\alpha\beta} \cdot \boldsymbol{\sigma}$, with $\mathbf{H}_{\alpha\beta} = \left( H_{x,\alpha\beta}, H_{y,\alpha\beta}, H_{z,\alpha\beta} \right)$ a vector of three complex numbers, and $\boldsymbol{\sigma} = \left( \sigma_x, \sigma_y, \sigma_z \right)$ a vector of the three Pauli matrices. The self-energies, the $\Gamma$-matrices, the retarded and lesser Green's functions, and the density matrix are expanded in an analogous form. The equations derived in the foregoing part of this chapter are therefore also valid for the non-collinear case, but one needs to remove the spin index from each equation and replace the appearing matrices with their non-collinear counterparts given in Eq. (24).

We now present the relations that allow us to evaluate the NEGF spin-transfer torque (STT), which corresponds to the torque that the flowing electrons exert on the local magnetic moments and which leads to current-induced time-dependent spin dynamics. Using the Liouville equation for a general density matrix in a non-orthogonal basis, $\partial\rho/\partial t = \frac{1}{i\hbar} \left( \Omega^{-1} H\rho - \rho H \Omega^{-1} \right)$, we obtain the relation:

$$\frac{\partial\bar\rho}{\partial t} = \frac{1}{i\hbar} [H, \rho] + \frac{1}{i\hbar} [W, \Omega], \tag{25}$$

where we have introduced the symmetrized density matrix:

$$\bar\rho = \frac{1}{2} \left( \rho\Omega + \Omega\rho \right). \tag{26}$$

Here $W = \frac{1}{2} \left[ \Omega^{-1} H\rho + \rho H \Omega^{-1} \right]$, which for the steady state corresponds to Eq. (13). The total number of electrons, $N_{\text{e,tot}} = 2 \sum_{\alpha=1}^{N_\beta} \bar\rho_{0,\alpha\alpha}$, and the total spin, $\mathbf{S}_{\text{tot}} = \hbar \sum_{\alpha=1}^{N_\beta} \bar{\boldsymbol{\rho}}_{\alpha\alpha}$, depend only on the diagonal elements of $\bar\rho$. We introduce the Mülliken-partitioned local charge in orbital $\alpha$ as $q_\alpha = 2e \, \bar\rho_{0,\alpha\alpha}$ and the local spin as $\mathbf{s}_\alpha = \hbar\bar{\boldsymbol{\rho}}_{\alpha\alpha}$, so that $\bar\rho_{\alpha\alpha} = \frac{1}{2e} q_\alpha \, 1_2 + \frac{1}{\hbar} \mathbf{s}_\alpha \cdot \boldsymbol{\sigma}$. Note that $q_\alpha$ and $\mathbf{s}_\alpha$ are always real. Thus, from Eq. (25) we derive two equations of motion:

$$\frac{\partial q_\alpha}{\partial t} = I_{0,\alpha}, \tag{27}$$

$$\frac{\partial \mathbf{s}_\alpha}{\partial t} = \mathbf{T}_{\mathbf{H},\alpha} + \frac{\hbar}{2e} \mathbf{I}_\alpha. \tag{28}$$

Here we have introduced the field-induced orbital torque

$$\mathbf{T}_{\mathbf{H},\alpha} = 2 \sum_{\beta=1}^{N_\beta} \text{Re} \left[ \mathbf{H}_{\alpha\beta} \times \boldsymbol{\rho}_{\beta\alpha} \right]; \tag{29}$$

the orbital charge current, $I_{0,\alpha} = \sum_{\beta=1}^{N_\beta} I_{0,\alpha\beta}$; and the orbital spin current, $\mathbf{I}_\alpha = \sum_{\beta=1}^{N_\beta} \mathbf{I}_{\alpha\beta}$. These include the bond current between any two pairs of orbitals $\alpha$ and $\beta$, $\mathbf{I}_{\alpha\beta} = I_{0\alpha\beta} \, 1_2 + \mathbf{I}_{\alpha\beta} \cdot \boldsymbol{\sigma}$, for which using Eq. (25) the charge-current component results to:

$$I_{0,\alpha\beta} = \frac{4e}{\hbar} \text{Im} \left[ H_{0,\alpha\beta} \rho_{0,\beta\alpha} - \Omega_{\alpha\beta} W_{0,\beta\alpha} + \mathbf{H}_{\alpha\beta} \cdot \boldsymbol{\rho}_{\beta\alpha} \right], \tag{30}$$

and the spin-current component results to:

$$\mathbf{I}_{\alpha\beta} = \frac{4e}{\hbar} \text{Im} \left[ \mathbf{H}_{\alpha\beta} \rho_{0,\beta\alpha} + H_{0,\alpha\beta} \boldsymbol{\rho}_{\beta\alpha} - \Omega_{\alpha\beta} \mathbf{W}_{\beta\alpha} \right]. \tag{31}$$

Note that the orbitals $\alpha$ and $\beta$ can be located both inside and outside of the EM, and that $\mathbf{I}_{\alpha\beta} = -\mathbf{I}_{\beta\alpha}$ and $I_{0,\alpha\beta} = -I_{0,\beta\alpha}$. We can interpret Eq. (27) as the change of local charge due to in- and outflow of electrons from the orbital $\alpha$, the first term in Eq. (28) as torque leading to a spin precession around the effective field originating from the Hamiltonian matrix and the second term as change of spin on orbital $\alpha$ due to inflow and outflow of spins. In the steady state $I_{0,\alpha} = 0$ and $\mathbf{T}_{\mathbf{H},\alpha} = -\frac{\hbar}{2e} \mathbf{I}_\alpha$, so that the term in Eq. (28) proportional to $\mathbf{I}_\alpha$ can be interpreted as a spin-transfer torque from the flowing electrons to the local magnetization. We denote this STT as $\mathbf{T}_{\text{STT},\alpha} = \frac{\hbar}{2e} \mathbf{I}_\alpha$. In the non-steady state, $\mathbf{I}_\alpha$ can also lead to a change of magnitude of the spin.

If we denote with $A$ an arbitrary interface to the current flow, then the total current across this interface, $I_A$, is obtained by summing up the bond currents between all the orbitals $\alpha_L$ on the left side of this interface and the ones on the right of this interface, $\alpha_R$:

$$I_A = \sum_{\alpha_L, \alpha_R} I_{\alpha_L \alpha_R}. \tag{32}$$

Charge conservation at the steady state implies that the charge current, $I_{A,0}$, across any interface $A$ perpendicular to the transport direction must give the same value. In fact it can be shown that the current obtained in this way is equivalent to the NEGF current in Eq. (14). In contrast, the spin currents $\mathbf{I}_A$ can vary for two different interfaces $A$ and $A'$, and the difference between $\mathbf{I}_A$ and $\mathbf{I}_{A'}$ satisfies the relation:

$$\sum_{\alpha \in (A, A')} \mathbf{T}_{\mathrm{STT}, \alpha} = \frac{\hbar}{2e} \left( \mathbf{I}_{A'} - \mathbf{I}_A \right), \tag{33}$$

where the sum on the left side of the equation runs over all orbitals enclosed between the interfaces $A$ and $A'$, so that it is equal to the total STT in the region between $A$ and $A'$. In typical MR heterostructures shown in Fig. 2a, b, d, one is interested in the torque acting on the right magnetic electrode, $\mathbf{T}_R$, since it is the one that can reverse the magnetization, while the magnetization of the left electrode is usually pinned to one direction. $\mathbf{T}_R$ is then proportional to the difference in the spin current deep inside the right electrode, $\mathbf{I}_R$, and the one in the middle of the spacer, $\mathbf{I}_M$, and the value is given by

$$\mathbf{T}_R = \frac{\hbar}{2e} \left( \mathbf{I}_R - \mathbf{I}_M \right). \tag{34}$$

Note that $\mathbf{I}_R$ needs to be evaluated at a distance from the interface that is larger than the transverse spin relaxation length (Slonczewski 2005).

We can also compute the local torque on each atom, where the Mülliken-partitioned local atomic spin at an atom with integer index $a$ is $\mathbf{S}_a = \sum_{\alpha \in a} \mathbf{s}_\alpha$. Here the sum over $\alpha$ runs over all orbitals that belong to atom $a$, and $\mathbf{S}_{\mathrm{tot}} = \sum_a \mathbf{S}_a$. In the steady state the STT acting on the atom, $\mathbf{T}_a$, is then given by

$$\mathbf{T}_a = \sum_{\alpha \in a} \mathbf{T}_{\mathrm{STT}, \alpha} = \frac{\hbar}{2e} \sum_{\alpha \in a} \mathbf{I}_\alpha. \tag{35}$$

To obtain $\mathbf{T}_a$ we can therefore calculate directly $\mathbf{T}_{\mathrm{STT}, \alpha} = \frac{\hbar}{2e} \mathbf{I}_\alpha$ through Eq. (31). Note that since the spin-current can be non-zero even without applied voltage, the torque defined in this way can be finite even at 0 bias voltage. If one wants to calculate only the torque induced by the application of a bias voltage, one can simply take the difference $\mathbf{T}_a(V_b) - \mathbf{T}_a(V_b = 0)$.

In Haney et al. (2007) an alternative approximate procedure for the calculation of $\mathbf{T}_{H, \alpha}$ is outlined, where one subdivides the full non-equilibrium electron density into a *condensate* part, $\rho_{\mathrm{cond}}$, which gives rise to the local moments at equilibrium, and a *transport* part due to the current flow, $\rho_{\mathrm{tr}}$, so that $\rho = \rho_{\mathrm{cond}} + \rho_{\mathrm{tr}}$. We note that the exact way of splitting $\rho$ into a condensate and a transport part is rather arbitrary, and the results can depend sensitively on this splitting, as shown in Xie et al. (2016). In a steady-state current-carrying condition, we choose the definition $\rho_{\mathrm{cond}} = \rho_{\mathrm{EM}}(V_b = 0)$, so that $\rho_{\mathrm{tr}}(V_b) = \rho_{\mathrm{EM}}(V_b) - \rho_{\mathrm{EM}}(V_b = 0)$. Within the LSDA, and in absence of spin-orbit interactions, we also have $\mathbf{H}[\rho]_{\alpha\beta} \propto \boldsymbol{\rho}_{\alpha\beta}$, where $\mathbf{H}[\rho]_{\alpha\beta}$ denotes the spin Hamiltonian obtained for the density matrix $\rho$. Assuming $\rho_{\mathrm{cond}} \gg \rho_{\mathrm{tr}}$ we have $\partial \rho_{\mathrm{cond}}/\partial t \gg \partial \rho_{\mathrm{tr}}/\partial t \approx 0$ and therefore $\partial \mathbf{s}_{\mathrm{tr}}/\partial t \approx 0$, and furthermore $\mathbf{H}[\rho_{\mathrm{cond}} + \rho_{\mathrm{tr}}]_{\alpha\beta} \approx \mathbf{H}[\rho_{\mathrm{cond}}]_{\alpha\beta} + \mathbf{H}[\rho_{\mathrm{tr}}]_{\alpha\beta}$. With Eqs. (28) and (35), this gives the following approximation for the STT:
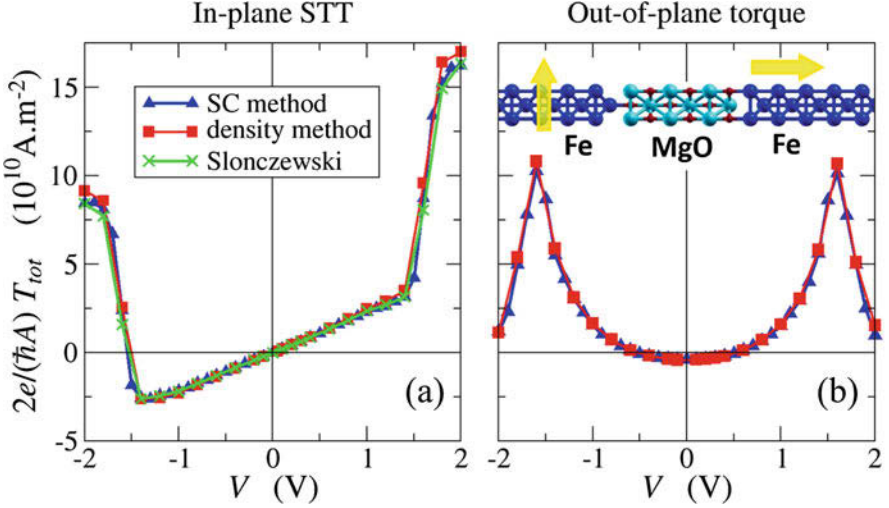
**Fig. 8** Comparison of the two non-collinear STT methods (density method and spin-current method) for the in-plane and out-of-plane torques in the right lead of an Fe/MgO/Fe(100) stack with magnetizations set at 90° (see inset), as function of the applied voltage. For the in-plane case, also the STT calculated using collinear-state transmissions, based on the equivalent-circuit magnetic tunnel junction model of Slonczewski (2005) developed for STT by Theodonis et al. (2006), is shown. Note that all finite-bias calculations here are not fully self-consistent, but are instead based on a non-self-consistent finite-bias calculation, where the potential drop across the barrier is assumed to be linear, which is a good approximation of the self-consistent potential (see Fig. 4)

$$\mathbf{T}_a \approx \frac{\partial \mathbf{s}_{a,\text{cond}}}{\partial t} \approx 2 \sum_{\alpha \in a} \sum_{\beta=1}^{N_\beta} \text{Re} \left[ \boldsymbol{\rho}_{\text{tr},\beta\alpha} \times \mathbf{H}[\rho_{\text{cond}}]_{\alpha\beta} \right]. \tag{36}$$

With Eq. (36) $\mathbf{T}_a$ can be evaluated by first performing a zero-bias calculation to obtain $\rho_{\text{EM}}(V_\text{b} = 0)$ and $\mathbf{H}[\rho_{\text{EM}}(V_\text{b} = 0)]_{\alpha\beta}$ and then a finite-bias calculation to obtain $\rho_{\text{EM}}(V_\text{b})$.

In Fig. 8 we show an example for the bias dependence of the total STT on the right lead of an Fe/MgO/Fe(100) magnetic tunnel junction (MTJ) calculated by the two methods above, namely, what we call the spin-current (SC) method based on Eq. (34) and what we denote as the density method (DM) of Eq. (36). We see a linear bias dependence of the in-plane (damping-like) STT for $|V| < 1.5$ V, above which a sharp increase is observed. The latter is due to the band edge of the $\Delta_1$ band for the minority spin in bcc Fe at about 1.5 eV above $E_\text{F}$ and the consequent sharp increase of the current in the AP alignment of the junction at these rather high bias voltages (see Fig. 3). The out-of-plane (field-like) torque is quadratic in this bias range, becoming larger than the in-plane STT at high bias voltages and then decaying sharply at around ±1.5 V. Importantly, we find an excellent agreement of the two presented methods over the considered wide range of bias voltages. The

results of these methods agree also very well with the Slonczewski's equivalent-circuit result for a symmetric MTJ, which is based on an approximate collinear spin-current model (Theodonis et al. 2006).

# 7 Linear-Response Spin-Transfer Torque

If we consider the linear-response regime, where only a small bias voltage, $\delta V_b$, is applied across the junction, then the transport part of the density matrix induced by $\delta V_b$ is

$$\rho_{tr}(\delta V_b) \approx \left. \frac{\partial \rho(V_b)}{\partial V_b} \right|_{V_b=0} \delta V_b. \tag{37}$$

Based on this relation, and using Eq. (36), we can evaluate the so-called spin-transfer torkance on a site $a$, $\boldsymbol{\tau}_a$, as

$$\boldsymbol{\tau}_a = \frac{\delta \mathbf{T}_a}{\delta V_b} \approx 2 \sum_{\alpha \in a} \sum_{\beta=1}^{N_\beta} \text{Re} \left[ \frac{\partial \boldsymbol{\rho}_{tr,\beta\alpha}}{\partial V_b} \times \mathbf{H}[\rho_{cond}]_{\alpha\beta} \right]. \tag{38}$$

Assuming that $G_{EM}(E)$ and $\Gamma_{L/R}(E)$ are slowly varying functions around $E_F$ and that the junction is approximately left-right symmetric (the two electrodes and the interfaces), resulting in a symmetric bias drop ($\pm V_b/2$), the derivative in above equation takes the form (in accordance with Mahfouzi and Nikolic (2013) for this special case):

$$\left. \frac{\partial \rho(V_b)}{\partial V_b} \right|_{V_b=0} = \frac{1}{4\pi} G_{EM}(E_F) \left[ \Gamma_L(E_F) - \Gamma_R(E_F) \right] G_{EM}^\dagger(E_F) \Big|_{V_b=0}. \tag{39}$$

In Fig. 9 an example calculation of a linear-response torkance is shown. In this case the electrodes are made of CuMnAs, an antiferromagnetic (AFM) material recently proposed for spintronic applications (Wadley et al. 2013, 2016). In its tetragonal phase, CuMnAs is a metallic AFM ($T_N \approx 480$ K) and can be grown epitaxially on GaP. The general advantages of AFMs are that they are immune to stray magnetic fields and that they are expected to have much faster switching times compared to ferromagnets. For CuMnAs grown on GaP, the magnetization is uniformly aligned within the individual Mn layers parallel to the interface and alternates antiferromagnetically from layer to layer as one moves away from the GaP interface (Fig. 9c, d). Since the spin-transport properties are dominated by the interface magnetization, reversal of the magnetization of the right electrode leads to a significant MR in this system (Fig. 9f). This particular equilibrium interface spin polarization is then found to give rise to staggered spin-transfer torkance in the AFM (Fig. 9e) (Stamenova et al. 2017), and the total torkance in the right electrode is non-zero despite the zero total magnetic moment. The magnetic state
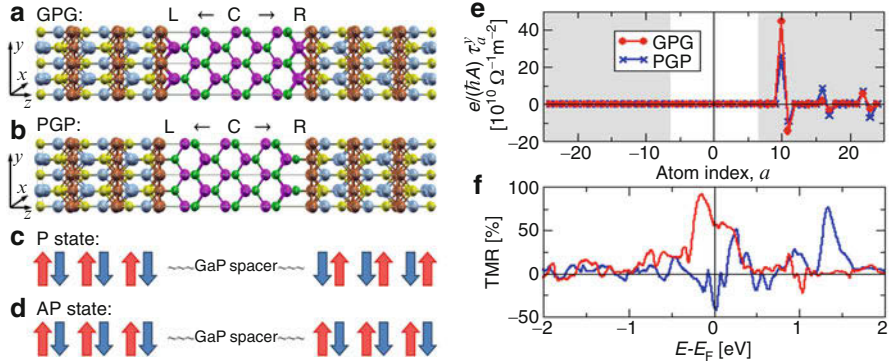
**Fig. 9** (**a,b**) Atomic structure of the CuMnAs/GaP/CuMnAs junctions; two possible terminations of the GaP spacer are considered: GPG ending with Ga layers and PGP ending with P; (**c,d**) definition of the P and AP spin states of the junction, arrows represent the spin of the corresponding Mn layer as per diagram above; (**e**) the in-plane spin torkance for a 90° spin misalignment, i.e., left lead polarized along $y$ and right lead along $x$; (**f**) energy-dependent linear-response TMR coefficient, which here is defined as $\text{TMR}(E) \equiv (T_{\text{P}}(E) - T_{\text{AP}}(E)) / (T_{\text{P}}(E) + T_{\text{AP}}(E))$. (Figure adapted from Stamenova et al. 2017)

of such heterostructures can therefore be read using the MR and switched using the torkance. We also note that, as shown in Fig. 9, the spin-transport properties are found to depend sensitively on the CuMnAs termination on GaP, but the staggered spin-transfer torkance is less sensitive to the details of the interface.

# 8 Time-Dependent Spin Dynamics: Current-Driven Domain Wall Motion

So far we have considered the charge and spin transport in the steady state. Time-dependent (TD) dynamical processes with non-collinear spin textures are typically evaluated in a multi-scale approach, where the NEGF+DFT STT steady-state results are used as parameters in a Landau-Lifshitz micromagnetic model (Xie et al. 2016; Ellis et al. 2017). The time evolution can in principle also be evaluated directly within NEGF+DFT; these calculations are however computationally highly demanding. Computations become significantly faster if instead of DFT one uses a simplified tight-binding (TB) model, where the Hamiltonian is based on a set of parameters instead of the fully first principles approach used in DFT. In Stamenova et al. (2005) a so-called $s - d$ model tight-binding Hamiltonian is used to describe a point contact, where the spatially varying localized spins, $\{\mathbf{S}_a\}$, are assumed to lie in a plane and are described by a one-dimensional set of classical angular coordinates $\boldsymbol{\phi} = \{\phi_a\}$, one for each site $a$. The corresponding Hamiltonian operator is written as

$$\hat{H}(\boldsymbol{\phi}) = \sum_{a,b} \left[ (H_{el})_{ab} + (V_{sd}(\boldsymbol{\phi}))_{ab} \right] \hat{c}_a^\dagger \hat{c}_b + V_{dd}(\boldsymbol{\phi}) \,, \tag{40}$$

where $a, b$ indicate the sites in the junction; $H_{el}$ is a single-band TB Hamiltonian matrix; $(V_{sd}(\boldsymbol{\phi}))_{ab} = -J_{sd}\mathbf{s}_a \cdot \mathbf{S}_b(\phi_b)\delta_{ab}$ corresponds to an on-site repulsion term; $\mathbf{S}_a(\phi_a) = (\sin(\phi_a), 0, \cos(\phi_a))$ are unit vectors representing the local atomic spins of uniform magnitude, $S$; and $\mathbf{s}_a = (0, 0, s_a) = (0, 0, \rho_{aa}^\uparrow - \rho_{aa}^\downarrow)/2$ is the expectation value of the on-site itinerant spin in the case of a two-spin-channel model (for ease of notation here and in the following, we define the spins as dimensionless). The equation includes an inter-site classical Heisenberg spin-spin interaction $V_{dd}(\{\phi_a\}) = -(J_{dd}/2)\sum_{a,b}\mathbf{S}_a(\phi_a)\cdot\mathbf{S}_b(\phi_b)$ with only nearest neighbor coupling. In this case the torques acting on the classical $d$-spins, defined through the Hellman-Feynman theorem, have an explicit form in the direction perpendicular to $\mathbf{S}_a$, allowing for spin-rotations energetics to be calculated, including transition barriers between different spin states in the junction. Within a NEGF steady-state spin-transport approach, this model has provided a valuable unequivocal numerical demonstration that the current-induced torques, which represent a special case of current-induced forces (CIFs), are nonconservative (Stamenova et al. 2005). The question of whether or not CIFs are conservative had been an open question in the field of electromigration and a subject of debate in the early 2000s (see, e.g., Di Ventra et al. 2004).

This model has been further extended beyond the two-spin channel model to fully non-collinear spins to investigate two almost reciprocal phenomena involving the dynamical interaction of electron transport and the underlying spin/magnetization texture in the time domain, i.e., the spin-motive force as a result of a domain wall (DW) rotation (Stamenova et al. 2008) on one hand and the current-induced DW motion (Stamenova and Sanvito 2010) on the other hand. The system of choice in the two cases was similar – a 1D chain of a few hundred localized spins forming a Néel-type DW. We will consider here in more detail the second case, where time-dependent steady-state transport is achieved by a single-parameter construction of an "ever-discharging" capacitor attached to the chain, analogous to the method of Sánchez et al. (2006). There it is also demonstrated that the steady state obtained in this method is equivalent to the NEGF steady state in the case where atomistic details of the coupling to the leads are neglected ($\Gamma_L(E) = \Gamma_R(E) = \Gamma$). An example of a possible TD transport scheme based on the NEGF is presented in McEniry et al. (2007). The $s-d$ spin Hamiltonian is now fully non-collinear, and the equation of motion for the localized spins reads (Stamenova and Sanvito 2010):

$$\dot{\mathbf{S}}_a = \frac{\mathbf{S}_a}{S} \times \left( J_{sd}\,\mathbf{s}_a + \gamma\,S\,\mathbf{B}(t) + J_{dd}\sum_b \mathbf{S}_b + 2J_{\text{ani}}\left(\mathbf{S}_a \cdot \hat{\mathbf{z}}\right)\hat{\mathbf{z}} \right) \,, \tag{41}$$

where $\mathbf{s}_a = \text{Tr}[\rho_{aa}\boldsymbol{\sigma}]/2$ is the instantaneous expectation value of the on-site itinerant spin and the electron density matrix is propagated according to a Liouville equation, including a "bias-maintaining" empirical damping term $\Gamma_b$ at the chain
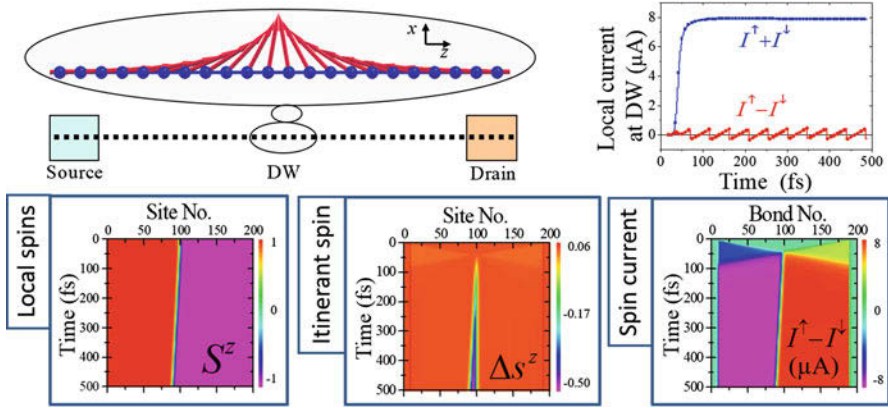
**Fig. 10** Top panels: a schematic of the DW-containing 1D conducting chain with open boundary conditions and the dynamic establishment of steady-state regime (showing charge and spin bond currents at the DW center). Bottom panels: 2D contour plots (as function of time and position along the chain) of spin-related observables as indicated. (Figure adapted from Stamenova and Sanvito 2010)

boundaries, which effectively generates a source and a drain, as illustrated in Fig. 10. Note that the STT component of this equation, corresponding to the STT presented in Eq. (36), is the first term in the brackets, STT = $J_{sd} \mathbf{S}_a \times \mathbf{s}_a/S$. For a suitable choice of $\Gamma_b$, a steady current-carrying state is established after a short transient time of a few tens of fs (Fig. 10b). It is found that during the steady state, the DM in the chain is moving at a practically constant velocity. This is only gently perturbed by the emission of spin waves (see Fig. 11). Under the steady current flow, the DW is developing some deformation, and two comparative cases have been examined – a very narrow and a very wide wall (on the length scale of the chain). Although the two walls develop similar velocities for a given current flow (the wider DW being somewhat faster) that scale linearly with the current, they undergo different levels of deformations (Fig. 11). This can be interpreted as a result of the different balance of current-induced torque contributions: the adiabatic STT (ASTT) and nonadiabatic torque (NAT). The ASTT is due to the adiabatic alignment of the itinerant electron spin to the local spin texture, and the NAT is a result of the lag in that spin alignment, which is expected to be more pronounced in the case of abrupt spin texture variation (e.g., a sharp DW). These are closely related to the driving mechanisms for the in-plane STT and the out-of-plane field-like torque in the MTJs (Brataas et al. 2012).

Alternatively, the current-driven DW motion can be studied by micromagnetic simulations based on the extended LLG equation incorporating STT terms, which have been shown to arise from the $s - d$ interaction by Zhang and Li (2004) in a diffusive semiclassical spin-transport model. An analogous to Eq. (41) fully classical atom-resolved LLG-like equation of motion, including the current-induced torques on the spin texture derived by Zhang and Li (2004), can be considered in the form:
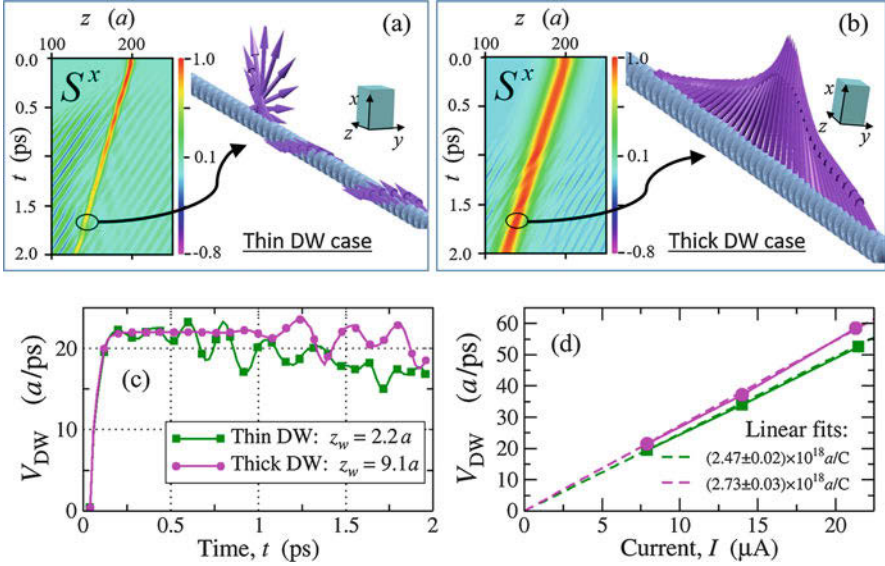
**Fig. 11** (**a,b**) Current-driven evolution of the $x$-components of the local spins is shown in contour plots and a snapshot of the domain wall (DW) spin structure at a late stage of the simulation, comparing two DWs with different thicknesses. (**c**) DW velocities of the two different walls are presented as function of time from the beginning of the simulation, and it can be seen that they are nearly constant once steady state is established after about 0.2 ps. (**d**) The average velocity of the two DWs as a function of the average steady-state current. Here $a$ is the lattice constant of the one-dimensional chain of atoms, $z_w$ is the width of the DW and C is the SI unit Coulomb. (Figure adapted from Stamenova and Sanvito 2010)

$$\dot{\mathbf{S}}_a = \mathbf{S}_a \times \left[ \frac{J_{dd}}{S} \sum_b \mathbf{S}_b + \frac{2J_{\text{ani}}}{S} \left( \mathbf{S}_a \cdot \hat{\mathbf{z}} \right) \hat{\mathbf{z}} + \alpha \, \partial_t \mathbf{S}_a - b_J \, \mathbf{S}_a \times \partial_z \mathbf{S}_a - c_J \, \partial_z \mathbf{S}_a \right],$$

(42)

where $b_J$ is the strength of the ASTT in units of velocity, $c_J$ is the strength of the NAT, and $\alpha$ is the Gilbert damping parameter. Restricting the degrees of freedom within the wall to two time-dependent parameters: the center and the width of the wall, Zhang and Li (2004) show that the ASTT drives a DW widening deformation and that the NAT is responsible for the terminal velocity. The atomistic simulations using Eq. (42) by Stamenova and Sanvito (2010) agree qualitatively in general, for example, the wall does not start moving with the NAT alone. Relaxing the DW shape rigidity condition however reveals additional features of the current-driven DW evolution. For instance, the NAT tends to give rise to the out-of-plane DW deformation and DW precession, similarly to what has been found by the quantum-classical steady-state simulations, comparing the thick and the thin walls (Fig. 11). Atomistic spin dynamics combined with first principles STTs from NEGF+DFT calculations have also recently been used to study the switching process of a free layer in an MTJ and investigate the effects of heating on the switching times (Ellis et al. 2017).

# 9 Conclusions

The NEGF+DFT formalism is a powerful tool to extract spin-dependent transport properties of nanoscale devices. It allows for quantitative and predictive calculations of spin currents, magnetoresistance, and spin-transfer torque, and in particular to evaluate the dependence of the device properties on atomic-scale features in the structure, which are often not directly accessible in experiment. As such it is now a well-established state-of-the-art tool for nanoscale device design. Despite the success so far, methodological developments are ongoing, in particular for the inclusion of strong correlation effects, of spin-phonon and spin-photon interactions, and for the time-dependent dynamic evolution.

# References

Alekhin A, Razdolski I, Ilin N, Meyburg JP, Diesing D, Roddatis V, Rungger I, Stamenova M, Sanvito S, Bovensiepen U, Melnikov A (2017) Femtosecond spin current pulses generated by the nonthermal spin-dependent Seebeck effect and interacting with ferromagnets in spin valves. Phys Rev Lett 119:017202–017207

Anisimov VI, Gunnarsson O (1991) Density-functional calculation of effective coulomb interactions in metals. Phys Rev B 43:7570

Baibich MN, Broto JM, Fert A, Van Dau FN, Petroff F, Etienne P, Creuzet G, Friederich A, Chazelas J (1988) Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. Phys Rev Lett 61:2472–2475

Binasch G, Grünberg P, Saurenbach F, Zinn W (1989) Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. Phys Rev B 39:4828–4830

Brataas A, Kent AD, Ohno H (2012) Current-induced torques in magnetic materials. Nat Mater 11(5):372–381

Butler WH, Zhang XG, Schulthess TC, MacLaren JM (2001) Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches. Phys Rev B 63:054416–054427

Büttiker M, Imry Y, Landauer R, Pinhas S (1985) Generalized many-channel conductance formula with application to small rings. Phys Rev B 31:6207–6215

Chioncel L, Morari C, Östlin A, Appelt WH, Droghetti A, Radonjić MM, Rungger I, Vitos L, Eckern U, Postnikov AV (2015) Transmission through correlated $Cu_nCoCu_n$ heterostructures. Phys Rev B 92:054431–054443

Datta S (1995) Electronic transport in mesoscopic systems. Cambridge University Press, Cambridge

Datta S (2005) Quantum transport: atom to transistor. Cambridge University Press, Cambridge

De Teresa JM, Barthélémy A, Fert A, Contour JP, Lyonnet R, Montaigne F, Seneor P, Vaurès A (1999) Inverse tunnel magnetoresistance in $Co/SrTiO_3/La_{0.7}Sr_{0.3}MnO_3$: new ideas on spin-polarized tunneling. Phys Rev Lett 82:4288–4291

Dolui K, Narayan A, Rungger I, Sanvito S (2014) Efficient spin injection and giant magnetoresistance in $Fe/MoS_2/Fe$ junctions. Phys Rev B 90:041401–041405(R)

Droghetti A, Rungger I (2017) Quantum transport simulation scheme including strong correlations and its application to organic radicals adsorbed on gold. Phys Rev B 95:085131

Ellis MOA, Stamenova M, Sanvito S (2017) Multiscale modeling of current-induced switching in magnetic tunnel junctions using ab initio spin-transfer torques. Phys Rev B 96(22):224410

Ghosh A (2017) Nanoelectronics: a molecular view. World Scientific Publishing Company, Singapore

Haney PM, Waldron D, Duine RA, Núñez AS, Guo H, MacDonald AH (2007) Current-induced order parameter dynamics: microscopic theory applied to Co/Cu/Co spin valves. Phys Rev B 76(2):024404

Himmetoglu B, Floris A, de Gironcoli S, Cococcioni M (2014) Hubbard-corrected dft energy functionals: the LDA+U description of correlated systems. Int J Quant Chem 114(1):14–49

Jacob D (2015) Towards a full ab initio theory of strong electronic correlations in nanoscale devices. J Phys Condens Matter 27(24):245606

Jutong N, Rungger I, Schuster C, Eckern U, Sanvito S, Schwingenschloegl U (2012) Electronic transport through EuO spin-filter tunnel junctions. Phys Rev B 86:205310–205319

Kotliar G, Savrasov SY, Haule K, Oudovenko VS, Parcollet O, Marianetti CA (2006) Electronic structure calculations with dynamical mean-field theory. Rev Mod Phys 78:865–951

Landauer R (1957) Spatial variation of currents and fields due to localized scatterers in metallic conduction. IBM J Res Dev 1:223–231

López Sancho MP, López Sancho JM, Rubio J (1984) Highly convergent schemes for the calculation of bulk and surface Green functions. J Phys F Met Phys 14:1205

Mahfouzi F, Nikolic BK (2013) How to construct the proper gauge-invariant density matrix in steady-state nonequilibrium: applications to spin-transfer and spin-orbit torques. SPIN 3(2):133002

McEniry EJ, Bowler DR, Dundas D, Horsfield AP, Sánchez CG, Todorov TN (2007) Dynamical simulation of inelastic quantum transport. J Phys Condens Matter 19(19):196201

Miyake T, Aryasetiawan F, Imada M (2009) Ab initio procedure for constructing effective models of correlated materials with entangled band structure. Phys Rev B 80:155134

Ness H, Dash LK, Godby RW (2010) Generalization and applicability of the Landauer formula for nonequilibrium current in the presence of interactions. Phys Rev B 82:085426

Parkin SSP, Kaiser C, Panchula A, Rice PM, Hughes B, Samant M, Yang SH (2004) Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. Nat Mater 3:862–867

Pemmaraju CD, Archer T, Sánchez-Portal D, Sanvito S (2007) Atomic-orbital-based approximate self-interaction correction scheme for molecules and solids. Phys Rev B 75:045101

Perdew JP, Levy M (1997) Comment on "significance of the highest occupied Kohn-Sham eigenvalue". Phys Rev B 56:16021–16028

Perdew JP, Zunger A (1981) Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B 23:5048

Pertsova A, Canali CM, Pederson MR, Rungger I, Sanvito S (2015) Electronic transport as a driver for self-interaction-corrected methods. Adv Atomic Mol Opt Phys 64:29–86

Rocha A, Garcìa-Suarez V, Bailey S, Lambert C, Ferrer J, Sanvito S (2005) Towards molecular spintronics. Nat Mater 4:335

Rocha A, Garcia-Suarez V, Bailey S, Lambert C, Ferrer J, Sanvito S (2006) Spin and molecular electronics in atomically generated orbital landscapes. Phys Rev B 73:085414–085433

Rungger I, Sanvito S (2008) Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition. Phys Rev B 78:035407–035419

Rungger I, Mryasov O, Sanvito S (2009) Resonant electronic states and I-V curves of Fe/MgO/Fe(100) tunnel junctions. Phys Rev B 79:094414–094418

Sánchez CG, Stamenova M, Sanvito S, Bowler DR, Horsfield AP, Todorov TN (2006) Molecular conduction: do time-dependent simulations tell you more than the Landauer approach? J Chem Phys 124(21):214708

Slonczewski JC (2005) Currents, torques, and polarization factors in magnetic tunnel junctions. Phys Rev B 71(2):024411–024420

Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, Sánchez-Portal D (2002) The SIESTA method for ab initio order-N materials simulation. J Phys Condens Matter 14(11):2745–2779

Souza AdM, Rungger I, Pontes RB, Rocha AR, da Silva AJR, Schwingenschloegl U, Sanvito S (2014) Stretching of BDT-gold molecular junctions: thiol or thiolate termination? Nanoscale 6:14495

Stamenova M, Sanvito S (2010) Atomistic spin-dynamics. In: Narlikar AV, Fu YY (eds) Oxford handbook of nanoscience and technology, vol 1. Oxford University Press, New York

Stamenova M, Sanvito S, Todorov TN (2005) Current-driven magnetic rearrangements in spin-polarized point contacts. Phys Rev B 72(13):134407

Stamenova M, Todorov TN, Sanvito S (2008) Newtonian origin of the spin motive force in ferromagnetic atomic wires. Phys Rev B 77(5):054439

Stamenova M, Mohebbi R, Seyed-Yazdi J, Rungger I, Sanvito S (2017) First-principles spin-transfer torque in CuMnAs|GaP|CuMnAs junctions. Phys Rev B 95(6):060403

Stefanucci G, van Leeuwen R (2013) Nonequilibrium many-body theory of quantum systems: a modern introduction. Cambridge University Press, Cambridge

Strange M, Rostgaard C, Häkkinen H, Thygesen KS (2011) Self-consistent GW calculations of electronic transport in thiol- and amine-linked molecular junctions. Phys Rev B 84:115108–115120

Theodonis I, Kioussis N, Kalitsov A, Chshiev M, Butler WH (2006) Anomalous bias dependence of spin torque in magnetic tunnel junctions. Phys Rev Lett 97(23):237205

Di Ventra M, Chen YC, Todorov TN (2004) Are current-induced forces conservative? Phys Rev Lett 92(17):176803

von Barth U, Hedin L (1972) A local exchange-correlation potential for the spin polarized case: I. J Phys C Solid State Phys 5(13):1629–1642

Wadley P, Novák V, Campion R, Rinaldi C, Martí X, Reichlová H, Železný J, Gazquez J, Roldan M, Varela M, Khalyavin D, Langridge S, Kriegner D, Máca F, Mašek J, Bertacco R, Holý V, Rushforth A, Edmonds K, Gallagher B, Foxon C, Wunderlich J, Jungwirth T (2013) Tetragonal phase of epitaxial room-temperature antiferromagnet CuMnAs. Nat Commun 4:2322

Wadley P, Howells B, Elezny J, Andrews C, Hills V, Campion RP, Novak V, Olejnik K, Maccherozzi F, Dhesi SS, Martin SY, Wagner T, Wunderlich J, Freimuth F, Mokrousov Y, Kune J, Chauhan JS, Grzybowski MJ, Rushforth AW, Edmonds KW, Gallagher BL, Jungwirth T (2016) Electrical switching of an antiferromagnet. Science 351(6273):587–590

Xie Y, Rungger I, Stamenova M, Sanvito S, Munira K, Ghosh AW (2016) Spin transfer torque: a multiscale picture. In: Atulasimha J, Bandyopadhyay S (eds) Nanomagnetic and spintronic devices for energy-efficient memory and computing. Wiley, Chichester, p 91

Yuasa S, Nagahama T, Fukushima A, Suzuki Y, Ando K (2004) Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. Nat Mater 3: 868–871

Zhang S, Li Z (2004) Roles of nonequilibrium conduction electrons on the magnetization dynamics of ferromagnets. Phys Rev Lett 93(12):127204

# Spintronics in Micromagnetics

# 45

Claas Abert

# Contents

## Abstract

The micromagnetic model has proven to be a reliable tool for the description of a variety of magnetic materials and systems. Classical micromagnetics describes the equilibrium and dynamics of magnetization configuration under the influence of magnetic fields and other material-specific energy contributions such as the exchange interaction. With the rise of spintronics, the interaction of

C. Abert (✉)
Christian Doppler Laboratory for Advanced Magnetic Sensing and Materials, Faculty of Physics, University of Vienna, Vienna, Austria
e-mail: claas.abert@univie.ac.at

spin-polarized currents with the magnetization has gained a lot of interest. This chapter aims to give an overview over extensions of the micromagnetic model for the description of spintronics effects.

# 1 Introduction

The micromagnetic model was originally developed as an analytical tool for the description of domain walls in ferromagnetic material (Brown 1963). In contrast to atomistic or ab initio descriptions of magnetic systems, it follows a mean-field-like approach where the discrete magnetic moments are approximated by a continuous magnetization field $\mathbf{M}(\mathbf{x})$. However, in contrast to Maxwell's equations, the magnetization field in micromagnetics is assumed to have a constant modulus $|\mathbf{M}| = M_s \mathbf{m}$ with $\mathbf{m}$ being a unit vector field and $M_s$ being a material constant called the saturation magnetization. This fundamental assumption is based on the fact that the strong exchange interaction in ferromagnetic materials results in a near-parallel orientation of neighboring spins. Assuming a constant spin density thus yields a saturated field when averaging over small spin ensembles. In contrast to the macroscopic description of the magnetization in Maxwell's equations where the magnetization is averaged over multiple domains, the magnetization field in micro-magnetics spatially resolves not only domains but also the structure of domain walls.

Besides extrinsic energy contributions like the Zeeman energy due to external fields, the micromagnetic model accounts for a number of intrinsic energy contributions like the exchange energy which is the root of ferromagnetism. Since some of the intrinsic energy contributions have a quantum mechanical origin, micromagnetism is often referred to as semiclassical theory.

The accurate description of magnetic domains makes the micromagnetic theory a valuable tool for the investigation of magnetic hysteresis. Since the response time of a magnetic system is much shorter than the field evolution in typical hard-magnetic applications, hysteretic properties are usually computed by direct energy minimization; see, e.g., Fischbacher et al. (2017). For other purposes, the magnetization dynamics need to be resolved, e.g., in order to investigate the switching times in storage applications; see, e.g., Lee and Dieny (2006). The micromagnetic model, both static and dynamic, is defined in terms of partial differential equations and can be solved analytically only for simple limiting cases. With the rise of high-performance computing, the numerical solution of the micro-magnetic model has become an invaluable tool for the investigation of magnetic systems. However, classical micromagnetics as introduced in Brown (1963) does not describe the interaction of spin-polarized currents. These interactions that are usually summarized as spintronics effects have recently gained a lot of interest in the magnetic community.

This chapter is supposed to give an overview over existing extensions to the micromagnetic model for the description of spintronics effects. It summarizes the review article (Abert 2019) which the interested reader is referred to for a more detailed introduction to micromagnetics, spintronics, and numerical methods.

## 2 Energetics of a Ferromagnet

The total free energy of a system with respect to its magnetization is key to static and dynamic micromagnetics. Depending on the material of the magnetic system, the total energy is composed of multiple contributions. According to classical electrodynamics, every magnetic body $\Omega_m$ has a Zeeman energy given by

$$E^{\text{zee}} = -\mu_0 \int_{\Omega_m} M_s \mathbf{m} \cdot \mathbf{H}^{\text{zee}} \, d\mathbf{x} \tag{1}$$

with $\mathbf{H}^{\text{zee}}$ being the Zeeman field and $\mu_0$ being the vacuum permeability. An essential energy contribution for micromagnetics is the exchange energy which is responsible for ferromagnetism. As mentioned in the preceding section, the micromagnetic assumption of a saturated magnetization field is valid only for systems with strong local ordering forces. This requirement is met by the exchange interaction that penalizes nonparallel alignment of neighboring spins. The exchange energy can be derived as the continuum limit of a Heisenberg Hamiltonian $H = \mathbf{s}_1 \cdot \mathbf{s}_2$ assuming a symmetric lattice structure and is given by

$$E^{\text{ex}} = \int_{\Omega_m} A \sum_{i,j} \left( \frac{\partial m_i}{\partial x_j} \right)^2 d\mathbf{x} = \int_{\Omega_m} A(\nabla \mathbf{m})^2 \, d\mathbf{x} \tag{2}$$

where $A$ is called the exchange constant and $(\nabla \mathbf{m})^2 = \sum_{i,j} (\partial m_i / \partial x_j)^2$ is to be understood as a Frobenius inner product. Although usually derived from the Heisenberg model, this energy expression accurately describes the exchange energy in various materials including band magnets; see Hubert and Schäfer (1998).

Another important intrinsic energy contribution is the demagnetization energy, also referred to as magnetostatic energy. It describes the energy of the magnetization in its self-generated dipolar field. According to classical electrodynamics, the demagnetization field can be expressed as negative gradient of a scalar potential $\mathbf{H}^{\text{dem}} = -\nabla u$ with the potential $u$ being the solution to the partial differential equation

$$\nabla \cdot (-\nabla u + \mathbf{M}) = 0 \quad \text{in} \quad \mathbb{R}^3 \tag{3}$$

with open boundary conditions

$$u(\mathbf{x}) = \mathcal{O}(1/|\mathbf{x}|) \text{ for } |\mathbf{x}| \to \infty, \tag{4}$$

i.e., the potential decays to zero at infinite distance from the magnetic region $\Omega_m$. The solution to this problem can be stated in terms of the integral equation

$$u(\mathbf{x}) = -\frac{1}{4\pi} \left[ \int_{\Omega_m} \frac{\nabla' \cdot \mathbf{M}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \, d\mathbf{x}' - \int_{\partial \Omega_m} \frac{\mathbf{M}(\mathbf{x}') \cdot \mathbf{n}}{|\mathbf{x} - \mathbf{x}'|} \, d\mathbf{s}' \right], \tag{5}$$

see Jackson (1999). The energy connected to the demagnetization field is given by

$$E^{\text{dem}} = -\frac{\mu_0}{2} \int_{\Omega_t} \mathbf{M} \cdot \mathbf{H}^{\text{dem}} \, d\mathbf{x} \qquad (6)$$

where the factor $1/2$ accounts for the quadratic dependence of the energy on the magnetization $\mathbf{M}$.

Depending on the material, the magnetization preferably aligns to certain axes in a magnetic body called easy axes. This effect originates from anisotropic spin-orbit coupling either due to an anisotropic crystal structure or due to lattice deformation at material interfaces; see Hubert and Schäfer (1998). In the case of a single easy axis, this anisotropy energy is given by

$$E^{\text{aniu}} = -\int_{\Omega_m} [K_{u1}(\mathbf{m} \cdot \mathbf{e}_u)^2 + K_{u2}(\mathbf{m} \cdot \mathbf{e}_u)^4] \, d\mathbf{x} \qquad (7)$$

where $\mathbf{e}_u$ is a unit vector parallel to the easy axis and $K_{u1}$ and $K_{u2}$ are the scalar anisotropy constants. The expression for this uniaxial anisotropy is derived in a phenomenological fashion by considering all terms that comply with the twofold symmetry up to fourth order in $\mathbf{m}$ (Brown 1963). Materials with cubic lattice symmetry such as iron or nickel exhibit three easy axes $\mathbf{e}_i$ which are pairwise orthogonal $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$. Like for the uniaxial anisotropy, the expression for the cubic anisotropy energy is developed as series in magnetization components along the easy axes up to sixth order

$$E^{\text{anic}} = \int_{\Omega} [K_{c1}(m_1^2 m_2^2 + m_2^2 m_3^2 + m_3^2 m_1^2) + K_{c2} m_1^2 m_2^2 m_3^2] \, d\mathbf{x} \qquad (8)$$

where $m_i = \mathbf{e}_i \cdot \mathbf{m}$ is the projection of the magnetization $\mathbf{m}$ on the anisotropy axis $\mathbf{e}_i$. Only contributions compatible with the cubic symmetry are considered.

Both, Eqs. (7) and (8) hold for magnetic anisotropies in bulk material. If magnetic anisotropy is caused by lattice deformation at material interfaces, the energy depends on the magnetization configuration $\mathbf{m}$ at this interface only. The energy for such a surface anisotropy is obtained by similar expressions as (7) and (8). However, instead of integrating over the magnetic volume $\Omega_m$, the integration in this case has to be carried out over the respective interface $\partial \Omega_m$ only.

While the presented energy contributions form the basis for the description of various system, several additional energy contributions may be required for specific systems. Recently, antisymmetric exchange interaction has been extensively studied, since materials with antisymmetric exchange enable the formation of stable magnetic skyrmions (Bogdanov and Rößler 2001). Other energy contributions arise from the existence of eddy currents (see Torres et al. 2003 and Hrkac et al. 2005), Oersted fields (see Hertel 2001), or magnetostriction (see Fabian and Heider 1996 and Shu et al. 2004).

## 3 Static Micromagnetics

Static micromagnetics deals with the investigation of stable magnetization configurations. An important application area of this theory is the computation of hysteresis loops of permanent magnets (Schrefl et al. 1994). The prerequisite for a stable configuration is that it minimizes the energy while complying with the micromagnetic unit-sphere constraint

$$\min E(\mathbf{m}) \quad \text{subject to} \quad |\mathbf{m}(\mathbf{x})| = 1. \tag{9}$$

Using a Lagrange-multiplier ansatz, this problem can be solved by minimizing

$$E_\lambda(\mathbf{m}, \lambda) = E(\mathbf{m}) + \int_{\Omega_m} \lambda(\mathbf{x})\big(|\mathbf{m}|^2 - 1\big)\, \mathrm{d}\mathbf{x}, \tag{10}$$

with respect to the magnetization $\mathbf{m}$ and the Lagrange-multiplier field $\lambda$. Variation of the magnetization yields

$$\delta E_\lambda(\{\mathbf{m}, \lambda\}, \mathbf{v}_m) == \int_{\Omega_m} \frac{\delta E}{\delta \mathbf{m}} \cdot \mathbf{v}_m\, \mathrm{d}\mathbf{x} + \int_{\partial\Omega_m} \mathbf{B}(\mathbf{m}) \cdot \mathbf{v}_m\, \mathrm{d}\mathbf{s} + 2\int_{\Omega_m} \lambda \mathbf{m} \cdot \mathbf{v}_m\, \mathrm{d}\mathbf{x}. \tag{11}$$

where the boundary condition $\mathbf{B}$ depends on the choice of energy contributions (Abert 2019). Since this condition has to hold for arbitrary $\mathbf{v}_m$, the minimum-energy magnetization has to fulfill $\delta E / \delta \mathbf{m} = -2\lambda \mathbf{m}$ which is equivalent to Brown's condition

$$\frac{\delta E}{\delta \mathbf{m}} \times \mathbf{m} = 0; \tag{12}$$

see Brown (1963).

Testing (11) with functions that are defined on the boundary only $\mathbf{v}(\Omega_m \setminus \partial\Omega_m) = 0$ yields the additional boundary condition

$$\mathbf{B}(\mathbf{m}) = 0. \tag{13}$$

Variation of (10) with respect to the Lagrange-multiplier field $\lambda$ accounts for the micromagnetic unit-sphere constraint.

Depending on the energy contribution, the variational derivative $\delta E / \delta \mathbf{m}$ gives rise to different boundary conditions $\mathbf{B}$. The derivative of the Zeeman energy is given by

$$\frac{\delta E^{\text{zee}}}{\delta \mathbf{m}} = -\mu_0 M_s \mathbf{H}^{\text{zee}} \tag{14}$$

while no boundary condition **B** is introduced. The derivative of the exchange energy is given by

$$\frac{\delta E^{\text{ex}}}{\delta \mathbf{m}} = -2 \boldsymbol{\nabla} \cdot (A \boldsymbol{\nabla} \mathbf{m}) \tag{15}$$

with the additional Neumann boundary condition $\mathbf{B} = 2A \partial \mathbf{m}/\partial \mathbf{n} = 0$ that arises through integration by parts which is necessary to transform the variational differential to the variational derivative. Similar to the Zeeman energy, the derivative of the demagnetization energy reads

$$\frac{\delta E^{\text{dem}}}{\delta \mathbf{m}} = -\mu_0 M_s \mathbf{H}^{\text{dem}} \tag{16}$$

with no additional boundary condition. The derivative of the uniaxial and cubic anisotropy is defined by

$$\frac{\delta E^{\text{aniu}}}{\delta \mathbf{m}} = -2K_{\text{u}1} \mathbf{e}_u (\mathbf{e}_u \cdot \mathbf{m}) - 4K_{\text{u}2} \mathbf{e}_u (\mathbf{e}_u \cdot \mathbf{m})^3 \tag{17}$$

and

$$\frac{\delta E^{\text{anic}}}{\delta \mathbf{m}} = 2K_{\text{c}1} \begin{pmatrix} m_1 m_2^2 + m_1 m_3^2 \\ m_2 m_3^2 + m_2 m_1^2 \\ m_3 m_1^2 + m_3 m_2^2 \end{pmatrix} + 2K_{\text{c}2} \begin{pmatrix} m_1 m_2^2 m_3^2 \\ m_1^2 m_2 m_3^2 \\ m_1^2 m_2^2 m_3 \end{pmatrix} \tag{18}$$

respectively.

## 4 Dynamic Micromagnetics

The micromagnetic theory provides models not only for the investigation of stable magnetization configurations but also for the investigation of time- and space-resolved magnetization dynamics. The central equation for the magnetization dynamics is the Landau-Lifshitz equation originally introduced in Landau and Lifshitz (1935) and reformulated as Landau-Lifshitz-Gilbert equation in Gilbert (1955)

$$\partial_t \mathbf{m} = -\gamma \mathbf{m} \times \mathbf{H}^{\text{eff}} + \alpha \mathbf{m} \times \partial_t \mathbf{m} \tag{19}$$

where $\gamma = \mu_0 \gamma_e \approx 2.2128 \times 10^5$ m/As is the reduced gyromagnetic ratio and $\alpha > 0$ is the dimensionless Gilbert damping constant. The so-called effective field $\mathbf{H}^{\text{eff}}$ accounts for all energy contributions of the magnetic system and reads

$$\mathbf{H}^{\text{eff}} = -\frac{1}{\mu_0 M_s} \frac{\delta E}{\delta \mathbf{m}}. \tag{20}$$
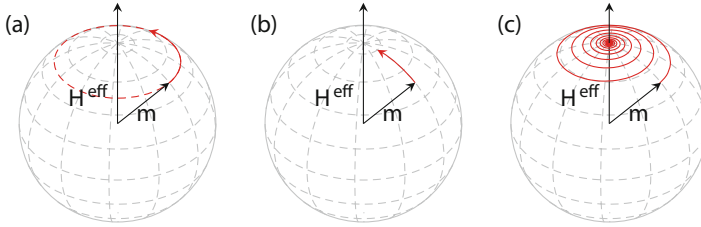
**Fig. 1** Magnetization dynamics as described by the Landau-Lifshitz-Gilbert (LLG) equation. (**a**) Precessional term. (**b**) Dissipative term. (**c**) Combined motion with precessional and damping contribution

With the choice of an appropriate Lagrangian, the LLG can be derived by variational calculus (see Wegrowe and Ciornei 2012), which explains the variational definition of the effective field. As for static micromagnetics, depending on the considered energy contributions, an addition boundary condition $\mathbf{B} = 0$ applies (see Sect. 3).

The semi-implicit formulation (19) introduced by Gilbert can be transformed into an explicit form by inserting the complete right-hand side of (19) into $\partial_t \mathbf{m}$ on the right-hand side of (19) and applying basic vector algebra

$$\partial_t \mathbf{m} = -\frac{\gamma}{1 + \alpha^2} \mathbf{m} \times \mathbf{H}^{\text{eff}} - \frac{\alpha\gamma}{1 + \alpha^2} \mathbf{m} \times (\mathbf{m} \times \mathbf{H}^{\text{eff}}). \tag{21}$$

This form is similar to the original formulation of the LLG proposed in Landau and Lifshitz (1935). Both in the implicit Gilbert form (19) and in the explicit form (21), the right-hand side of the LLG is composed of two terms. The first one is responsible for a precessional motion, and the second one, called damping term, accounts for energy dissipation due to eddy currents, lattice excitations, and other mechanisms (see Fig. 1).

## 5    Spintronics in Micromagnetics

The term spintronics was coined in the 1980s with the discovery of the giant magnetoresistance (GMR) by Fert and Grünberg and denotes spin-dependent interactions of conducting electrons with solid-state devices. In contrast to electronics which exploits the electron's charge, spintronics adds extra degrees of freedom and is believed to enable a variety of novel devices particularly in the area of storage and sensing.

In the continuous micromagnetic theory, the spin polarization of the conducting electrons is described by a vector field $\mathbf{p}$ that is bidirectionally coupled to both the magnetization $\mathbf{m}$ and the charge current $\mathbf{j}_e$. When passing magnetized regions, the conducting electrons gain polarization depending on the charge current and magnetization configuration. Reversely, the spin polarization affects the charge current by introducing a magnetization-dependent resistance, e.g., GMR. Moreover,

a spin-polarized current is able to manipulate a magnetization configuration by exerting a torque. Depending on the origin of the spin polarization, this torque is either referred to as spin-transfer torque or spin-orbit torque.

## 5.1    Spin-Transfer Torque in Multilayers

A possible device that generates and exploits spin-polarized currents is a magnetic multilayer. Depending on the magnetization configuration in the different magnetic layers, scattering processes produce spin-polarized currents which results in spin-transfer torque. A typical spin-transfer torque device has two magnetic layers, separated either with a nonmagnetic conducting layer or a tunnel junction and contacted via nonmagnetic leads. Figure 2 illustrates the dominating scattering processes in such a device for parallel and antiparallel magnetization configurations.

Following the electron flow, the first scattering process for any magnetization configuration occurs at the first nonmagnetic-magnetic interface where the minority electrons are scattered with a higher probability than the majority electrons. Consequently the electrons entering the second magnetic layer carry the polarization of the first magnetic layer FM1, i.e., FM1 acts as a spin filter. For the antiparallel magnetization configuration, this means that the majority of electrons entering FM2 are polarized antiparallel to the FM2 magnetization although some of these electrons are scattered at the interface of FM2 (see Fig. 2a). As a result, spin torque occurs in FM2 and leads to magnetization switching if a critical current is exceeded. The polarized electrons scattered at the FM2 interface reenter FM1 and stabilize its magnetization.

For the parallel configuration, the situation is depicted in Fig. 2b. As for the antiparallel configuration, FM1 acts as a spin filter which leads to a stabilization of FM2. The minority electrons scattered at the FM2 interface reenter FM1 and exert a torque that can switch the FM1 magnetization if a critical current is exceeded.

This torque mechanism was first predicted in Slonczewski (1996), Berkov et al. (1993), and Waintal et al. (2000). Possible applications for this effect include
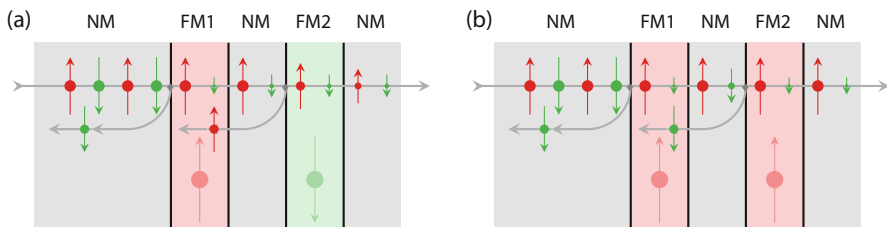


**Fig. 2** Dominating scattering processes in multilayer structures subject to a perpendicular charge current. FM1 and FM2 are ferromagnetic layers, while NM are nonmagnetic conducting layers. (**a**) FM2 is destabilized by polarized electrons coming from FM1. FM1 is stabilized by polarized electronics scattered back from FM2. (**b**) FM2 is stabilized by polarized electrons coming from FM1. FM1 is destabilized by polarized electrons scattered back from FM2

spin-transfer torque magnetic random-access memory (STT MRAM) as described in Huai (2008) and Worledge et al. (2011) and spin-torque oscillators (STO) as described in Houssameddine et al. (2007) and Kim (2012). A comprehensive theoretical overview over spin-transfer torque is given in Ralph and Stiles (2008).

In the context of micromagnetics, the model introduced in Slonczewski (2002) is widely used for the description of spin-transfer torque. The original model applies to a macrospin description of the magnetic layers, i.e., every layer is described by a single spin. One of the two magnetic layers is assumed to have a fixed magnetization denoted as $\mathbf{p}$ and is called polarizing layer since it acts as a spin polarizer. The other layer is referred to as free layer, and its magnetization $\mathbf{m}$ is subject to the spin torque generated by the polarizing layer. According to the Slonczewski model, the motion of the free-layer magnetization $\mathbf{m}$ is described by an extended LLG

$$\partial_t \mathbf{m} = -\gamma \mathbf{m} \times \mathbf{H}^{\text{eff}} + \alpha \mathbf{m} \times \partial_t \mathbf{m} + \mathbf{T} \tag{22}$$

with the torque $\mathbf{T} = \mathbf{T}_{\text{damp}} + \mathbf{T}_{\text{field}}$ being the sum of a so-called damping-like torque $\mathbf{T}_{\text{damp}}$ and a field-like torque $\mathbf{T}_{\text{field}}$. The model of Slonczewski defines these contributions as

$$\mathbf{T}_{\text{damp}} = \eta_{\text{damp}}(\vartheta) \frac{j_e \gamma \hbar}{2e\mu_0 M_s} \mathbf{m} \times (\mathbf{m} \times \mathbf{p}) \tag{23}$$

$$\mathbf{T}_{\text{field}} = \eta_{\text{field}}(\vartheta) \frac{j_e \gamma \hbar}{2e\mu_0 M_s} \mathbf{m} \times \mathbf{p} \tag{24}$$

where $\eta_{\text{damp}}$ and $\eta_{\text{field}}$ are dimensionless functions describing the angular dependence of the spin-torque efficiency with $\vartheta$ being the angle between the polarization $\mathbf{p}$ and the free-layer magnetization $\mathbf{m}$.

The field-like torque (24) is similar to the torque generated by an external field. It causes a damped precessional motion of the free-layer magnetization with $\alpha$ being the scaling of the damping. The damping-like torque also causes a damped precessional motion. However, in contrast to the field-like torque, the precessional motion scales with $\alpha$, whereas the dissipative motion is independent from $\alpha$. This fact becomes apparent when considering a system with vanishing damping constant $\alpha = 0$. In this case the extended LLG reads

$$\partial_t \mathbf{m} = -\gamma \mathbf{m} \times \left[ \mathbf{H}^{\text{eff}} - \frac{j_e \hbar}{2e\mu_0 M_s} \eta_{\text{field}}\, \mathbf{p} \right] - \gamma \mathbf{m} \times \left( \mathbf{m} \times \left[ \frac{j_e \hbar}{2e\mu_0 M_s} \eta_{\text{damp}}\, \mathbf{p} \right] \right). \tag{25}$$

For this limiting case, the precessional motion is exclusively generated by the field-like torque, while the damping-like motion is exclusively generated by the damping-like torque, which shows the unique features of spin-transfer torque.

In the original work by Slonczewski, the angular dependence $\eta(\vartheta)$ was derived for a symmetric device, namely, a device with identical polarizing layer and free layer, and reads

$$\eta(\vartheta) = \frac{P\Gamma}{(\Gamma + 1) + (\Gamma - 1)\cos(\vartheta)} \tag{26}$$

where the dimensionless $P$ is a measure for the polarization of the electrons and $\Gamma$ is a dimensionless parameters depending on geometry and material composition of the complete device. Expression (26) is used for the damping-like torque as well as for the field-like torque but requires individual parameters $P$ and $\Gamma$ for the different torques. For the accurate description of nonsymmetric devices, a generalized model was introduced in Xiao et al. (2005) which reads

$$\eta(\vartheta) = \frac{q^+}{A + B\cos(\vartheta)} + \frac{q^-}{A - B\cos(\vartheta)}. \tag{27}$$

As for the symmetric case, the expression is used for both the damping-like torque and the field-like torque with a different set of dimensionless parameters $q^+$, $q^-$, $A$, and $B$ that depend on geometry and materials in a nontrivial fashion.

While the polarizing layer in the model of Slonczewski is generally considered to have a fixed magnetization, the model can also be used to describe the bidirectional coupling of two magnetic layers. This can be achieved by a second set of model parameters and considering a second torque term with polarizing layer and free layer interchanged (see, e.g., Rowlands and Krivorotov 2012).

Since the original model of Slonczewski uses the macrospin approach, another generalization is required in order to account for spatially varying magnetization configurations. Magnetic layers with small sizes below the single-domain limit are dominated by the exchange coupling and hence are almost homogeneously magnetized. While the macrospin approach is a valid assumption for these layers, it fails for larger layers that exhibit magnetic domains and hence cannot be represented by a single spin. For lateral magnetization inhomogeneities, the macrospin approach can be generalized as depicted in Fig. 3. Instead of describing the magnetization in both magnetic layers with a single spin, they are described with fields as usual in micromagnetics. The Slonczewski model is then applied pointwise to nearest sites across the interface. Perpendicular inhomogeneities are handled by considering the spin-transfer to be a pure surface effect. That means that both the
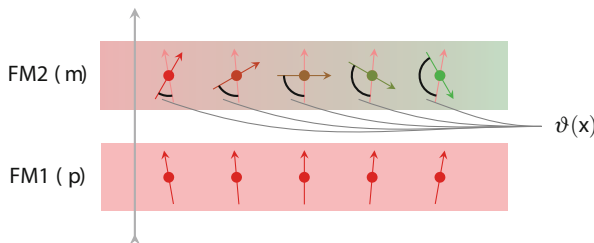


**Fig. 3** Application of the Slonczewski model for laterally varying magnetization configurations. The model is applied pointwise by considering the angle $\vartheta$ between the projected magnetization **p** of the polarizing layer FM1 and the magnetization **m** in the free layer

generation of spin polarization and the action of the spin torque are restricted to the magnetic-nonmagnetic interfaces. While this approach provides reasonable results for a number of applications, it neglect various diffusion processes which leads to inaccuracies. A comprehensive overview over Slonczewski-like spin-torque models is given in Berkov and Miltat (2008).

## 5.2    Spin-Transfer Torque in Continuous Media

The mechanism of spin-transfer torque is not restricted to magnetic multilayer structures but also appears in continuous magnetization configurations. In this case, magnetic domains take the place of distinct magnetic layers in the generation of spin-transfer torque. This process is depicted in a simplified fashion in Fig. 4. The electrons passing a magnetic domain are polarized parallel to its magnetization. When entering a region of magnetization gradient, namely, a domain wall, they carry this polarization in regions of nonparallel magnetization which leads to a spin torque. This process facilitates the current driven motion of domain walls in the direction of the electron flow. In contrast to field-driven domain wall motion, this process allows in theory for the coherent motion of complete domain structures. This characteristic is aimed to be exploited in the magnetic racetrack memory that was proposed in Parkin et al. (2008).

The model of Slonczewski is obviously not suited for the description of this spin-torque effect since it requires the existence of two distinct magnetization layers. For the description of spin torque in continuous magnetization configurations, a micromagnetic model was proposed in Zhang and Li (2004). This Zhang-Li model introduces a torque term given by

$$\mathbf{T} = -b\,\mathbf{m} \times [\mathbf{m} \times (\mathbf{j}_{\mathrm{e}} \cdot \nabla)\mathbf{m}] - b\xi\mathbf{m} \times (\mathbf{j}_{\mathrm{e}} \cdot \nabla)\mathbf{m} \qquad (28)$$
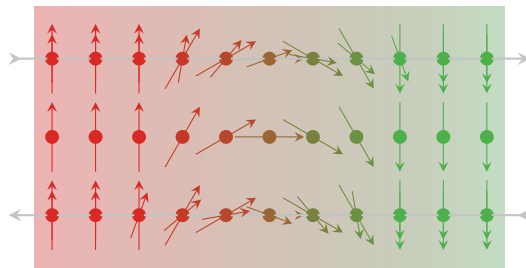


**Fig. 4** Spin-torque mechanism in a magnetic domain wall according to the model of Zhang and Li. The polarization of the magnetization is carried by the conducting electrons in their direction of motion. By neglecting diffusion, the Zhang-Li models consider this spin transport only in a local sense, i.e., depending on the local gradient of the magnetization. While the top row illustrates an electron flow from the left to the right, the bottom row illustrates the inverse direction. The magnetization is depicted by desaturated spins

where the dimensionless $\xi$ is called the degree of nonadiabacity and $b$ is defined as

$$b = \frac{\beta \mu_{\mathrm{B}}}{e M_{\mathrm{s}} (1 + \xi^2)} \tag{29}$$

with $\beta$ being a dimensionless polarization rate, $\mu_{\mathrm{B}}$ being the Bohr magneton, and $e$ being the elementary charge.

The Zhang-Li model describes spin-transfer torque purely in terms of local magnetization gradients, hence neglecting any diffusion processes which would lead to nonlocal effects. While this assumption is reasonable for materials with small characteristic diffusion lengths, it leads to inaccurate results for other materials, especially in the case of strongly inhomogeneous magnetization configurations. Moreover, the Zhang-Li model is not able to describe the nonlocal effect of spin torque in multilayers.

## 5.3    Spin-Diffusion

Neither the model of Slonczewski introduced in Sect. 5.1 nor the model of Zhang and Li introduced in Sect. 5.2 provide a comprehensive description of spin torque. Both models are restricted to certain material systems, and both models neglect diffusion effects to a certain extent. A more general spin-torque model is the three-dimensional generalization of the spin-diffusion model introduced in Zhang et al. (2002). In the spin-diffusion model, the polarization of the charge current is described by the so-called spin accumulation $\mathbf{s}$. The spin accumulation describes the deviation of a charge current's spin polarization from the equilibrium polarization at vanishing current. In nonmagnetic materials, the spin accumulation $\mathbf{s}$ equals the polarization $\mathbf{p}$. However, in magnetic materials, conducting electrons may contribute to the total magnetization and hence have polarization $\mathbf{p}^0$ even at vanishing current $\mathbf{j}_{\mathrm{e}} = 0$. The spin accumulation is then defined as $\mathbf{s} = \mathbf{p} - \mathbf{p}^0$.

Since the equilibrium spin polarization does not induce any torque on the magnetization, the spin torque in the spin-diffusion model can be defined in terms of the spin accumulation and reads

$$\mathbf{T} = -\frac{J}{\hbar M_{\mathrm{s}}} \mathbf{m} \times \mathbf{s} \tag{30}$$

where $J$ denotes the coupling strength of the spin accumulation $\mathbf{s}$ and the magnetization $\mathbf{m}$. The spin accumulation itself is subject to the equation of motion

$$\partial_t \mathbf{s} = -\boldsymbol{\nabla} \cdot \tilde{\mathbf{j}}_{\mathrm{s}} - \frac{\mathbf{s}}{\tau_{\mathrm{sf}}} - J \frac{\mathbf{s} \times \mathbf{m}}{\hbar} \tag{31}$$

where $\tau_{\mathrm{sf}}$ is a material constant and denotes the spin-flip relaxation time. Moreover, the spin-diffusion model defines a matrix-valued spin current $\tilde{\mathbf{j}}_{\mathrm{s}}$ and a vector-valued charge current $\mathbf{j}_{\mathrm{e}}$ as

$$\tilde{\mathbf{j}}_s = 2C_0\beta\frac{\mu_B}{e}\mathbf{m}\otimes\nabla u - 2D_0\nabla\mathbf{s} \tag{32}$$

$$\mathbf{j}_e = -2C_0\nabla u + 2D_0\beta'\frac{e}{\mu_B}(\nabla\mathbf{s})^T\mathbf{m} \tag{33}$$

where $u$ is the electric potential, $\mu_B$ is the Bohr magneton, and $e$ is the elementary charge. The remaining variables are the diffusion constant $D_0$, the dimensionless polarization parameter $\beta$ that is a measure for the polarization properties of magnetic materials, and $C_0$ which is connected to the electric conductivity $\sigma$ by the relation $\sigma = 2C_0$. It is important to note, that the spin accumulation is usually not only solved in the magnetic region $\Omega_m$ like the LLG, but in the complete device region $\Omega$ which might include nonmagnetic regions. That said, all material parameters might vary spatially and are hence described by scalar fields rather than constants.

In order to compute the spin accumulation, the spin current needs to be computed according to (32). However, the distribution of the electric potential $u$ is usually not known for a given system, and hence the definition of the electric current $\mathbf{j}_e$ needs to be taken into account, too. If $\mathbf{j}_e$ is given, the spin current $\tilde{\mathbf{j}}_s$ can be determined by inserting (33) into (32) via the potential $u$ which yields

$$\tilde{\mathbf{j}}_s = -\frac{\beta\mu_B}{e}\mathbf{m}\otimes\mathbf{j}_e - 2D_0\left(\nabla\mathbf{s} - \beta\beta'\mathbf{m}\otimes\left[(\nabla\mathbf{s})^T\mathbf{m}\right]\right). \tag{34}$$

Inserting into (31) yields the dynamics of the spin accumulation. The resulting magnetization dynamics can be obtained by coupling the solution of (31) to the solution of the extended LLG (22).

The problem with this approach, however, is the different time scales of spin-accumulation dynamics and magnetization dynamics. In most systems the spin accumulation relaxes two orders of magnitude faster than the magnetization (see Zhang and Li 2004). In order to determine the magnetization dynamics, it is hence sufficient to consider the spin accumulation $\mathbf{s}$ to instantaneously relax on magnetization changes, i.e., $\partial_t\mathbf{s} = 0$, which simplifies (31) to

$$\nabla\cdot\tilde{\mathbf{j}}_s + \frac{\mathbf{s}}{\tau_{sf}} + J\frac{\mathbf{s}\times\mathbf{m}}{\hbar} = 0 \quad\text{in}\quad \Omega. \tag{35}$$

In this quasistatic approach, the spin accumulation is obtained by the solution of a linear partial differential equation of second order in $\mathbf{s}$. Typical solutions of $\mathbf{s}$ for magnetic multilayers as well as domain walls are depicted in Fig. 5. The treatment of $\mathbf{s}$ in this approach becomes similar to the treatment of effective-field contributions that exclusively depend on the magnetization rather than the time.

So far, the computation of the spin accumulation $\mathbf{s}$ requires the knowledge of the charge current $\mathbf{j}_e$. However, the charge current itself is subject to spintronics effects, e.g., the giant magnetoresistance (GMR) effect predicts that the resistance of a spintronics device depends on its magnetization configuration. In order to solve
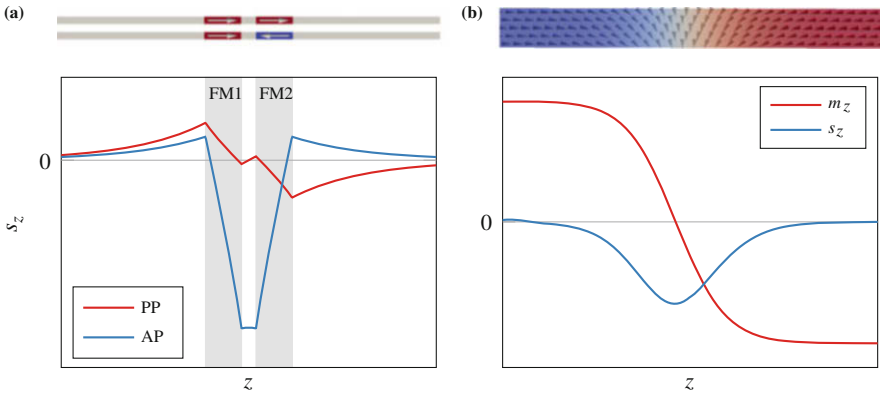
**Fig. 5** Spatially resolved spin accumulation **s** for typical magnetization configurations. Only the dominated component in shown. (**a**) Spin accumulation $s_z$ for magnetic multilayer with parallel and antiparallel magnetization configuration, respectively. (**b**) Magnetic domain wall. The magnetization $m_z$ is plotted against the spin accumulation $s_z$

the spin-diffusion model in a self-consistent fashion, the source equation for the spin current (36) needs to be complemented by a source equation for the charge current $\mathbf{j}_e$. Since magnetic devices usually do not contain sources of charge currents, the source equation for $\mathbf{j}_e$ is naturally given by the continuity equation

$$\nabla \cdot \mathbf{j}_e = 0 \quad \text{in} \quad \Omega. \tag{36}$$

The source Eqs. (36) and (35) together with the current definitions (33) and (32) yield a coupled system of partial differential equations with the solution variables being both the spin accumulation **s** and the electric potential $u$. For this self-consistent approach, applying potentials and currents to a magnetic system is implemented by boundary conditions.

## 5.4   Boundary Conditions for the Spin-Diffusion Model

The self-consistent spin-diffusion model is a second-order partial-differential equation in both the electric potential $u$ and as the spin accumulation **s**. Hence, boundary conditions are required in order to retrieve a unique solution. A typical multilayer system for the application of the spin-diffusion model is depicted in Fig. 6. In order to account for an electric current flowing perpendicular through the layer system, appropriate boundary conditions for the electric potential $u$ have to be defined on the contact interfaces $\Gamma_1$ and $\Gamma_2$. While a Dirichlet boundary condition, e.g., $u = u_0$ on $\Gamma_1$, prescribes a defined potential on the respective contact interface, a Neumann condition can be applied in order to account for a defined current

**Fig. 6** Typical circular
magnetic multilayer
consisting of two magnetic
layers FM1 and FM2,
separated by a nonmagnetic
layer NM and sandwiched by
two nonmagnetic leads with
top and bottom contacts on
the surfaces $\Gamma_1$ and $\Gamma_2$
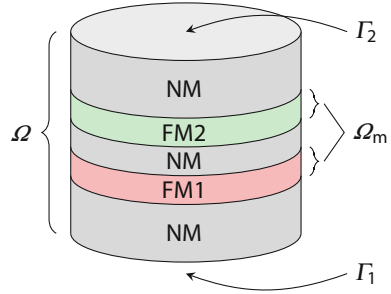


inflow. For instance, setting a constant current inflow $j_0$ on the second contact $\Gamma_2$ is
achieved by requiring

$$\mathbf{j}_e \cdot \mathbf{n} = -2 \left[ C_0 \nabla u + D_0 \beta' \frac{e}{\mu_B} \left[ (\nabla \mathbf{s})^T \mathbf{m} \right] \right] \cdot \mathbf{n} = j_0 \quad \text{on} \quad \Gamma_2 \qquad (37)$$

where $\mathbf{n}$ denotes the outward pointing normal to $\Gamma_2$. The boundary conditions for $u$
are completed by requiring vanishing current inflow on all boundaries that do not
act as contacts.

The spin accumulation $\mathbf{s}$, as the second solution variable, is solved with homoge-
neous Neumann boundary conditions on the complete boundary

$$\nabla \mathbf{s} \cdot \mathbf{n} = 0 \quad \text{on} \quad \partial \Omega. \qquad (38)$$

Inserting this condition into the definition of the spin current (34) and multiplying
with the boundary normal $\mathbf{n}$ yields

$$\tilde{\mathbf{j}}_s \cdot \mathbf{n} = \beta \frac{\mu_B}{e} \mathbf{m}(\mathbf{j}_e \cdot \mathbf{n}) \qquad (39)$$

which is equivalent to a noflux condition on the spin current $\tilde{\mathbf{j}}_s \cdot \mathbf{n} = 0$ if the charge
current $\mathbf{j}_e$ enters the system $\Omega$ in nonmagnetic regions only. This noflux condition
itself is assumed physically reasonable for all parts of the boundary with no charge-
current flux. For the contact region on the boundary $\Gamma_1$ and $\Gamma_2$, the noflux condition
is considered a reasonable assumption if the thickness of the nonmagnetic lead
regions exceeds the spin-diffusion length. If this is not the case, the homogeneous
Neumann condition should be replaced by the Robin condition

$$\nabla \mathbf{s} \cdot \mathbf{n} + \frac{1}{\sqrt{2D_0 \tau_{sf}}} \mathbf{s} = 0 \qquad (40)$$

which takes the exponential decay of $\mathbf{s}$ in nonmagnetic regions into account.

## 5.5    Extension to the Spin-Diffusion Model

The spin-diffusion model, introduced in the preceding sections, describes the generation of spin currents and spin accumulation due to polarizing magnetic materials. However, it was predicted in Dyakonov and Perel (1971) that spin currents can also be generated in nonmagnetic materials subject to charge currents due to spin-orbit coupling. This effect, referred to as spin-Hall effect, has its origin in the polarization-dependent deflection of the conducting electrons either due to material impurities as described in Hirsch (1999) or due to intrinsic asymmetries in the material as described in Murakami et al. (2003) and Sinova et al. (2004). While the spin-Hall effect enables the conversion from charge current into spin currents, the spin-orbit coupling also causes the inverse spin-Hall effect that enables the conversions of spin currents into charge currents. Both effects are schematically visualized in Fig. 7.

The integration of these effects with the spin-diffusion model is achieved by extending the original current definition (33) and (32) according to

$$j'_{e,i} = j_{e,i} + \varepsilon_{ijk}\theta_{SH}\frac{e}{\mu_B}j_{s,jk} \tag{41}$$

$$j'_{s,ij} = j_{s,ij} - \varepsilon_{ijk}\theta_{SH}\frac{\mu_B}{e}j_{e,k} \tag{42}$$

where index notation is used and $\varepsilon_{ijk}$ is the Levi-Civita tensor. Here, $\mathbf{j}_e$ and $\tilde{\mathbf{j}}_s$ are the original current definitions, and $\mathbf{j}'_e$ and $\tilde{\mathbf{j}}'_s$ are the current definitions including the spin-orbit couplings. Inserting $\mathbf{j}'_e$ and $\tilde{\mathbf{j}}'_s$ into the source Eqs. (36) and (35) instead of $\mathbf{j}_e$ and $\tilde{\mathbf{j}}_s$ then yields the spin-diffusion model including spin-orbit interactions. The dimensionless material constant $\theta_{SH}$ is called spin-Hall angle and describes the efficiency of the current conversion.
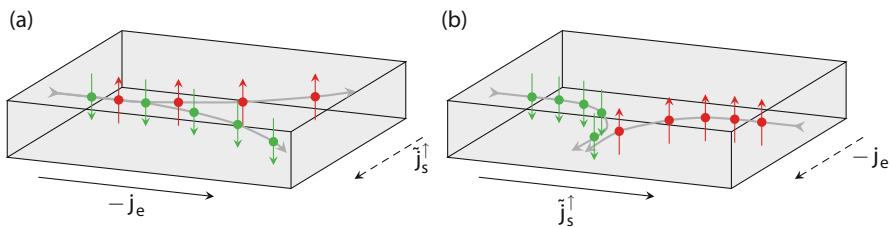


**Fig. 7** Spin current-charge current conversion due to spin-orbit torque-induced spin-dependent deflection processes. (**a**) Spin-Hall effect: A nonpolarized charge current $\mathbf{j}_e$ is converted to a pure spin current $\tilde{\mathbf{j}}_s$. (**b**) Inverse spin-Hall effect: A pure spin current $\tilde{\mathbf{j}}_s$ is converted to a charge current $\mathbf{j}_e$

Another extension of the spin-diffusion model is the addition of a spin-dephasing term, which was proposed in Akosa et al. (2015) and Haney et al. (2013). With this term the source equation for the spin current is extended to

$$\partial_t \mathbf{s} = -\boldsymbol{\nabla} \cdot \tilde{\mathbf{j}}_\mathrm{s} - \frac{\mathbf{s}}{\tau_\mathrm{sf}} - J \frac{\mathbf{s} \times \mathbf{m}}{\hbar} - \frac{\mathbf{m} \times (\mathbf{s} \times \mathbf{m})}{\tau_\phi} \tag{43}$$

where $\tau_\phi$ is the spin-dephasing time.

## 5.6    Valet-Fert Model

A related model to the spin-diffusion model introduced in the preceding sections is the Valet-Fert diffusion model introduced in Valet and Fert (1993). While the original model only accounts for magnetic multilayers with collinear magnetization, a three-dimensional generalization to noncollinear configurations is provided in Niimi et al. (2012). Like the spin-diffusion model from Sect. 5.3, the Valet-Fert model introduces a spin accumulation $\mathbf{s}$ and a spin current $\tilde{\mathbf{j}}_\mathrm{s}$. However, in contrast to the spin-diffusion model, both the spin accumulation and the spin current in magnetic materials are assumed to be collinear to the magnetization

$$\mathbf{s} = s\,\mathbf{m} \tag{44}$$

$$\tilde{\mathbf{j}}_\mathrm{s} = \mathbf{m} \otimes \mathbf{j}_\mathrm{s}. \tag{45}$$

According to (30), this definition of $\mathbf{s}$ would lead to vanishing spin torque. However, in contrast to the spin-diffusion model, the Valet-Fert model allows for discontinuities of the electric potential $u$ and the spin accumulation $\mathbf{s}$ across nonmagnetic-magnetic interfaces. These discontinuities are described by jump conditions that explicitly depend on distinct interface properties like the spin-mixing conductance $g_{\uparrow\downarrow}$. Hence, a noncollinear spin accumulation can occur at interfaces allowing for an interface-based description of spin torque. A detailed discussion of the generalized Valet-Fert model can be found in Niimi et al. (2012).

## 5.7    Connecting the Spintronics Models

The preceding sections discuss several models for the description of spintronics effects. While the Slonczewski model and the Zhang-Li model focus on spin torque in specific systems (see Sects. 5.1 and 5.2), other models like the spin-diffusion model introduced in Sect. 5.3 cover a variety of effects. In the following the connections of the different models will be discussed in detail.

### 5.7.1 Slonczewski Model

The Slonczewski model is a spin-torque model specifically designed for the description of magnetic multilayer structures. While the existence of a damping-like and a field-like torque is a rather generic assumption, the angular dependence $\eta$ introduced by Slonczewski is specific to the model. The angular dependence (26) proposed in the original paper is derived for symmetric structures with two similar magnetic layers, while the more general expression (27) removes this restriction. Since the model of Slonczewski is well established for the description of magnetic multilayers, a more general model like the spin-diffusion model is expected to reproduce its results. In order to compare both models, two simple multilayer systems with homogeneous magnetization configurations in the magnetic layers are investigated with the spin-diffusion model. The first system is a symmetric multilayer with two similar 3 nm thick magnetic layers, separated with a 2 nm metallic spacer, with typical material parameters as given in Abert et al. (2018). The spin accumulation **s** and the resulting torque for various angles between free-layer magnetization and polarizing-layer magnetization are computed with the spin-diffusion model. Projection of the total torque yields the angular dependence $\eta$ of the damping-like torque that is depicted in Fig. 8a. Along with the spin-diffusion results, Fig. 8a shows a fit to the original Slonczewski-formula (26) which exhibits perfect agreement.

The second system under consideration is an asymmetric multilayer which differs from the symmetric system by a higher polarizing layer thickness of 5 nm. As shown in Fig. 8b, the angular dependence of this asymmetric system is insufficiently fitted by the original Slonczewski formula but shows excellent agreement with the general expression (27). By reproducing the angular dependence of the spin torque, the spin-diffusion model proves to incorporate the Slonczewski model while
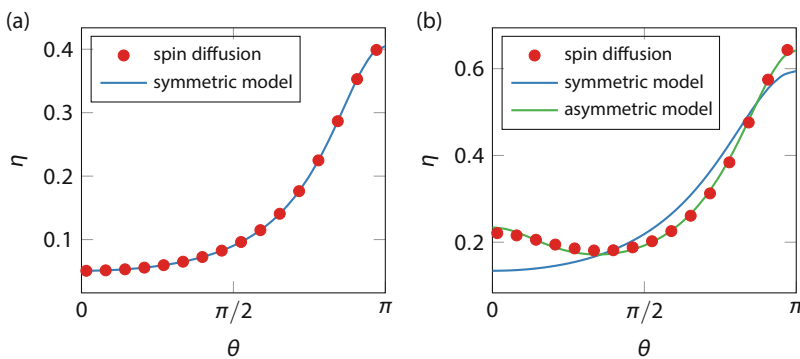


**Fig. 8** Spin-torque angular dependence $\eta$ as computed with the spin-diffusion model for homogeneous magnetization configurations in a magnetic multilayer with tilting angle $\vartheta$. The results are fitted with the Slonczewski formulas for symmetric and asymmetric structures, respectively. (**a**) Angular dependence $\eta$ for a symmetric multilayer with similar ferromagnetic layer thicknesses. (**b**) Angular dependence $\eta$ for an asymmetric multilayer

having a much broader application range including further spintronics effects. Moreover, the free parameters in the spin-diffusion model are material parameters with physical meaning in contrast to the free parameters of the Slonczewski-model that depend on the system as a whole in a nontrivial fashion.

### 5.7.2 Zhang-Li Model

Like the Slonczewski model, the Zhang-Li model discussed in Sect. 5.2 is a specialized model that describes the effect of spin torque in specific systems. It can be derived from the more general spin-diffusion model by neglecting diffusive effects, i.e., by setting $D_0 = 0$. Considering the definition of the spin current (34) and the source equation (35), this simplification yields

$$-\nabla \left( \frac{\beta \mu_B}{e} \mathbf{m} \otimes \mathbf{j}_e \right) + \frac{\mathbf{s}}{\tau_{sf}} + \frac{J}{\hbar} \mathbf{s} \times \mathbf{m} = \tag{46}$$

$$-\frac{\beta \mu_B}{e} (\mathbf{j}_e \cdot \nabla) \mathbf{m} + \frac{\mathbf{s}}{\tau_{sf}} - \frac{J}{\hbar} \mathbf{m} \times \mathbf{s} = 0. \tag{47}$$

Multiplication with $\mathbf{m}$ and $\mathbf{m} \times \mathbf{m}$, respectively, and inserting via $\mathbf{m} \times (\mathbf{m} \times \mathbf{s})$ result in the torque

$$\mathbf{T} = -\frac{J}{\hbar M_s} \mathbf{m} \times \mathbf{s} \tag{48}$$

$$= \frac{\beta \mu_B}{e M_s} \frac{1}{1 + \left( \frac{\hbar}{J \tau_{sf}} \right)^2} \left( \mathbf{m} \times [\mathbf{m} \times (\mathbf{j}_e \cdot \nabla) \mathbf{m}] + \frac{\hbar}{J \tau_{sf}} \mathbf{m} \times (\mathbf{j}_e \cdot \nabla) \mathbf{m} \right) \tag{49}$$

which exactly reproduces the spin torque of the Zhang-Li model (28) with the degree of nonadiabacity being defined as

$$\xi = \frac{\hbar}{J \tau_{sf}}. \tag{50}$$

While the calculation of the spin torque requires the solution of a linear system in the case of the spin-diffusion model, the same calculation reduces to the evaluation of expression (49) for the Zhang-Li model. However, this simplification restricts the application of the Zhang-Li model to systems with neglectable diffusive effects.

### 5.7.3 Valet-Fert Model

As discussed in Sect. 5.6, the Valet-Fert model is very similar to the spin-diffusion model introduced in Sect. 5.3. However, while the spin-diffusion model exclusively considers bulk material parameters for spin-transport properties, the Valet-Fert model introduces additional interface properties. These interface properties can be modelled in the spin-diffusion model by introducing thin effective interface layers. The respective material parameter mappings are discussed in detail in Abert (2019).

## 5.8 Beyond the Spin-Diffusion Model

The spin-diffusion model introduced in Sect. 5.3 accounts for various spintronics effects. However, it is certainly not a complete model for spintronics. While some spintronics effects like spin pumping, anomalous Hall effect, and inplane GMR are in principle compatible to the spin-diffusion model and could be introduced by suitable extensions, the main restriction of this model is its focus on diffusive transport. An important class of spintronics devices comprises magnetic tunnel junctions that are not well described by diffusive transport models. Various ab initio models for the description of magnetic tunnel junctions have been proposed; see, e.g., Mathon and Umerski (2001), Caffrey et al. (2011), and Butler et al. (2001). Integration of ab initio models with micromagnetics has been done (see Ellis et al. 2017) but is still a subject of ongoing research, partly due to the high computational complexity that is usually introduced by ab initio calculations.

## 6 Discretization

As a continuum theory, the micromagnetic model is defined in terms of partial differential equations. While analytical solutions to these equations can be derived for simple problem, the micromagnetic model is usually solved numerical. Suitable numerical algorithms have to account for various challenges that are particular to micromagnetic problems. Micromagnetic problems are multiscale problems. The demagnetization field is a global interaction which usually implies a computational complexity of $\mathcal{O}(n^2)$ when implemented naively. Various efficient methods for the demagnetization-field computation have been proposed in order to reduce the complexity to $\mathcal{O}(n \log n)$ or even $\mathcal{O}(n)$. The exchange field on the other hand is a short-range interaction that competes with the demagnetization field. While the computation of short-range fields is computationally cheap, the exchange field adds a high stiffness to the problem due to its second order in space. This high stiffness calls for stable time-integration schemes in order to solve the LLG. Another challenge of the micromagnetic model is its high degree of nonlinearity which leads to a complex energy landscape. This complexity leads to a high risk of missing local energy minima when computing hysteresis properties.

Among the existing discretization strategies, the most established methods used in micromagnetics are the finite-difference method and the finite-element method. A major difference of these approaches is the spatial discretization which is required to be a regular cuboid grid in the case of finite differences and may be an irregular, usually tetrahedral, mesh in the case of finite elements (see Fig. 9).

A detailed description of classical finite-difference micromagnetics is presented in Miltat and Donahue (2007). Several software packages implementing the finite-difference method were developed. One of the most popular packages is the open-source package OOMMF (see Donahue 1999), a multi-platform code running on central processing units (CPUs). Other CPU packages include the open-source
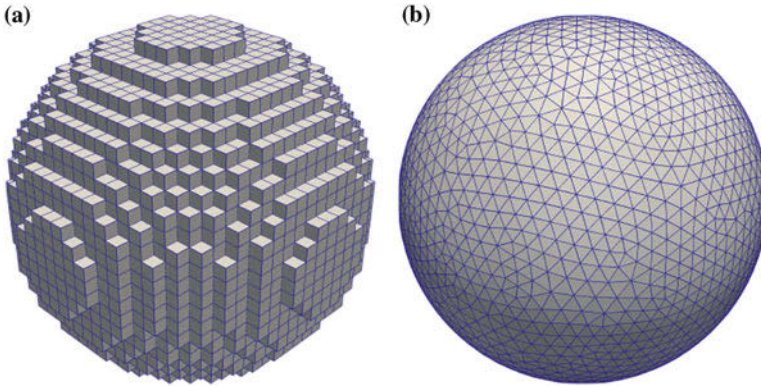
**Fig. 9** Spatial discretization of a sphere for different numerical methods. (**a**) Regular cuboid grid with 8217 cells as required for finite-difference micromagnetics. (**b**) Tetrahedral mesh with 7149 vertices as required for finite-element micromagnetics

software Fidimag (see Cortés-Ortuño et al. 2016) and the commercial software MicroMagus (see Berkov and Gorn 2007). A very basic finite-difference code written in Python was published in Abert et al. (2015).

The advent of multipurpose graphics processing units (GPGPUs) gave rise to a new class of simulation packages that take advantage of this highly parallel hardware. A very popular GPGPU package is MuMax3 (see Vansteenkiste et al. 2014). Alternatives are magnum.fd (see Selke et al. 2014) and the GPGPU extension to OOMMf (see Fu et al. 2016).

On overview over classical finite-element micromagnetics is given in Schrefl et al. (2007). Open-source finite-element packages include FinMag (see Bisotti et al. 2018) and Magpar (see Scholz 2010). Closed-source implementations include FEMME (see Suess and Schrefl 2018), FEELLGOOD (see Sturma et al. 2015), and magnum.fe (see Abert et al. 2013). GPGPU implementations include Tetramag (see Kakay et al. 2010) and Fastmag (see Chang et al. 2011).

Several of these simulation packages implement the spin-torque models of Slonczewski as well as of Zhang and Li. However, to the knowledge of the author, only the finite-element packages FEELLGOOD and magnum.fe implement the spin-diffusion model introduced in Sect. 5.3. A finite-difference implementation of this model, however, is described in García-Cervera and Wang (2007).

## 7    Conclusion

With the rise of high-performance computing, numerical micromagnetics has become an important tool for the development of magnetic devices. Micromagnetic simulations help to understand experimental measurements and guide the design of novel devices.

While classical micromagnetics relies on a well-approved theory that accurately describes numerous applications, the simulation of spintronics devices still poses challenges on the modelling procedure since various additional effects have to be considered. Simplified models like the spin-torque models of Slonczewski or Zhang and Li are computationally cheap and sufficiently accurate for a number of applications. However, the more general spin-diffusion model incorporates the simplified models and furthermore accounts for several additional spintronics effects. The accurate description of tunnel barriers in the context of micromagnetic simulations is still an active and important area of research.

# References

Abert C (2019) Micromagnetics and spintronics: models and numerical methods. print Eur Phys J B 92:120

Abert C, Exl L, Bruckner F, Drews A, Suess D (2013) magnum.fe: a micromagnetic finite-element simulation code based on fenics. J Magn Magn Mater 345:29–35

Abert C, Bruckner F, Vogler C, Windl R, Thanhoffer R, Suess D (2015) A full-fledged micromagnetic code in fewer than 70 lines of numpy. J Magn Magn Mater 387:13–18

Abert C, Bruckner F, Vogler C, Suess D (2018) Efficient micromagnetic modelling of spin-transfer torque and spin-orbit torque. AIP Adv 8(5):056008

Akosa CA, Kim WS, Bisig A, Kläui M, Lee KJ, Manchon A (2015) Role of spin diffusion in current-induced domain wall motion for disordered ferromagnets. Phys Rev B 91(9):094411

Berkov D, Gorn N (2007) MicroMagus–package for micromagnetic simulations. http://www.micromagus.de

Berkov DV, Miltat J (2008) Spin-torque driven magnetization dynamics: micromagnetic modeling. J Magn Magn Mater 320(7):1238–1259

Berkov D, Ramstöcck K, Hubert A (1993) Solving micromagnetic problems. towards an optimal numerical method. Phys Status Solidi (a) 137(1):207–225

Bisotti MA, Beg M, Wang W, Albert M, Chernyshenko D, Cortés-Ortuño D, Pepper RA, Vousden M, Carey R, Fuchs H, Johansen A, Balaban G, Breth L, Kluyver T, Fangohr H (2018) FinMag. https://github.com/fangohr/finmag

Bogdanov A, Rößler U (2001) Chiral symmetry breaking in magnetic thin films and multilayers. Phys Rev Lett 87(3):037203

Brown WF Jr (1963) Micromagnetics. Interscience Publisher, New York

Butler W, Zhang XG, Schulthess T, MacLaren J (2001) Spin-dependent tunneling conductance of fe| mgo| fe sandwiches. Phys Rev B 63(5):054416

Caffrey NM, Archer T, Rungger I, Sanvito S (2011) Prediction of large bias-dependent magnetoresistance in all-oxide magnetic tunnel junctions with a ferroelectric barrier. Phys Rev B 83(12):125409

Chang R, Li S, Lubarda M, Livshitz B, Lomakin V (2011) FastMag: fast micromagnetic simulator for complex magnetic structures. J Appl Phys 109(7):07D358

Cortés-Ortuño D, Wang W, Pepper R, Bisotti MA, Kluyver T, Vousden M, Fangohr H (2016) Fidimag v2.0. https://github.com/computationalmodelling/fidimag

Donahue MJ (1999) OOMMF user's guide, version 1.0. https://archive.org/stream/oommfusersguidev6376dona/oommfusersguidev6376dona_djvu.txt

Dyakonov M, Perel V (1971) Current-induced spin orientation of electrons in semiconductors. Phys Lett A 35(6):459–460

Ellis MO, Stamenova M, Sanvito S (2017) Multiscale modeling of current-induced switching in magnetic tunnel junctions using ab initio spin-transfer torques. Phys Rev B 96(22):224410

Fabian K, Heider F (1996) How to include magnetostriction in micromagnetic models of titanomagnetite grains. Geophys Res Lett 23(20):2839–2842

Fischbacher J, Kovacs A, Oezelt H, Schrefl T, Exl L, Fidler J, Suess D, Sakuma N, Yano M, Kato A, et al (2017) Nonlinear conjugate gradient methods in micromagnetics. AIP Adv 7(4): 045310

Fu S, Cui W, Hu M, Chang R, Donahue MJ, Lomakin V (2016) Finite-difference micromagnetic solvers with the object-oriented micromagnetic framework on graphics processing units. IEEE Trans Magn 52(4):1–9

García-Cervera CJ, Wang XP (2007) Spin-polarized currents in ferromagnetic multilayers. J Comput Phys 224(2):699–711

Gilbert TL (1955) A Lagrangian formulation of the gyromagnetic equation of the magnetic field. Phys Rev 100:1243

Haney PM, Lee HW, Lee KJ, Manchon A, Stiles MD (2013) Current induced torques and interfacial spin-orbit coupling: semiclassical modeling. Phys Rev B 87(17):174411

Hertel R (2001) Micromagnetic simulations of magnetostatically coupled nickel nanowires. J Appl Phys 90(11):5752–5758

Hirsch J (1999) Spin hall effect. Phys Rev Lett 83(9):1834

Houssameddine D, Ebels U, Delaët B, Rodmacq B, Firastrau I, Ponthenier F, Brunet M, Thirion C, Michel JP, Prejbeanu-Buda L et al (2007) Spin-torque oscillator using a perpendicular polarizer and a planar free layer. Nat Mater 6(6):447

Hrkac G, Kirschner M, Dorfbauer F, Suess D, Ertl O, Fidler J, Schrefl T (2005) Three-dimensional micromagnetic finite element simulations including eddy currents. J Appl Phys 97(10):10E311

Huai Y (2008) Spin-transfer torque MRAM (STT-MRAM): challenges and prospects. AAPPS Bull 18(6):33–40

Hubert A, Schäfer R (1998) Magnetic domains. Springer, Berlin

Jackson JD (1999) Classical electrodynamics, 3rd ed. Am J Phys 67(9):841. https://doi.org/10.1119/1.19136, http://link.aip.org/link/?AJP/67/841/2&Agg=doi

Kakay A, Westphal E, Hertel R (2010) Speedup of fem micromagnetic simulations with graphical processing units. IEEE Trans Magn 46(6):2303–2306

Kim JV (2012) Spin-torque oscillators. In: Solid state physics, vol 63. Elsevier, Academic Press, pp 217–294

Landau LD, Lifshitz EM (1935) On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. Physikalische Zeitschrift der Sowjetunion 8:153–169

Lee KJ, Dieny B (2006) Micromagnetic investigation of the dynamics of magnetization switching induced by a spin polarized current. Appl Phys Lett 88(13):132506

Mathon J, Umerski A (2001) Theory of tunneling magnetoresistance of an epitaxial fe/mgo/fe (001) junction. Phys Rev B 63(22):220403

Miltat JE, Donahue MJ (2007) Numerical micromagnetics: finite difference methods. In: Handbook of magnetism and advanced magnetic materials. Wiley, Chichester

Murakami S, Nagaosa N, Zhang SC (2003) Dissipationless quantum spin current at room temperature. Science 301(5638):1348–1351

Niimi Y, Kawanishi Y, Wei D, Deranlot C, Yang H, Chshiev M, Valet T, Fert A, Otani Y (2012) Giant spin hall effect induced by skew scattering from bismuth impurities inside thin film cubi alloys. Phys Rev Lett 109(15):156602

Parkin SS, Hayashi M, Thomas L (2008) Magnetic domain-wall racetrack memory. Science 320(5873):190–194

Ralph DC, Stiles MD (2008) Spin transfer torques. J Magn Magn Mater 320(7):1190–1216

Rowlands GE, Krivorotov IN (2012) Magnetization dynamics in a dual free-layer spin-torque nano-oscillator. Phys Rev B 86(9):094425

Scholz W (2010) MagPar. http://www.magpar.net/

Schrefl T, Fidler J, Kronmüller H (1994) Remanence and coercivity in isotropic nanocrystalline permanent magnets. Phys Rev B 49(9):6100

Schrefl T, Hrkac G, Bance S, Suess D, Ertl O, Fidler J (2007) Numerical methods in micromagnetics (finite element method). In: Handbook of magnetism and advanced magnetic materials. Wiley, Chichester

Selke G, Krüger B, Drews A, Abert C, Gerhardt T (2014) magnum.fd. https://github.com/micromagnetics/magnum.fd

Shu Y, Lin M, Wu K (2004) Micromagnetic modeling of magnetostrictive materials under intrinsic stress. Mech Mater 36(10):975–997

Sinova J, Culcer D, Niu Q, Sinitsyn N, Jungwirth T, MacDonald A (2004) Universal intrinsic spin hall effect. Phys Rev Lett 92(12):126603

Slonczewski JC (1996) Current-driven excitation of magnetic multilayers. J Magn Magn Mater 159(1–2):L1–L7

Slonczewski JC (2002) Currents and torques in metallic magnetic multilayers. J Magn Magn Mater 247(3):324–338

Sturma M, Toussaint JC, Gusakova D (2015) Geometry effects on magnetization dynamics in circular cross-section wires. J Appl Phys 117(24):243901

Suess D, Schrefl T (2018) FEMME. http://suessco.com/simulations/solutions/femme-software/

Torres L, Lopez-Diaz L, Martinez E, Alejos O (2003) Micromagnetic dynamic computations including eddy currents. IEEE Trans Magn 39(5):2498–2500

Valet T, Fert A (1993) Theory of the perpendicular magnetoresistance in magnetic multilayers. Phys Rev B 48(10):7099

Vansteenkiste A, Leliaert J, Dvornik M, Helsen M, Garcia-Sanchez F, Van Waeyenberge B (2014) The design and verification of mumax3. AIP Adv 4(10):107133

Waintal X, Myers EB, Brouwer PW, Ralph D (2000) Role of spin-dependent interface scattering in generating current-induced torques in magnetic multilayers. Phys Rev B 62(18):12317

Wegrowe JE, Ciornei MC (2012) Magnetization dynamics, gyromagnetic relation, and inertial effects. Am J Phys 80(7):607–611. https://doi.org/10.1119/1.4709188, http://link.aip.org/link/?AJP/80/607/1

Worledge D, Hu G, Abraham DW, Sun J, Trouilloud P, Nowak J, Brown S, Gaidis M, O'Sullivan E, Robertazzi R (2011) Spin torque switching of perpendicular Ta|CoFeB|MgO-based magnetic tunnel junctions. Appl Phys Lett 98(2):022501

Xiao J, Zangwill A, Stiles MD (2005) Macrospin models of spin transfer dynamics. Phys Rev B 72(1):014446

Zhang S, Li Z (2004) Roles of nonequilibrium conduction electrons on the magnetization dynamics of ferromagnets. Phys Rev Lett 93(12):127204

Zhang S, Levy P, Fert A (2002) Mechanisms of spin-polarized current-driven magnetization switching. Phys Rev Lett 88(23):236601

# Quantum Monte Carlo for Electronic Systems Containing *d* and *f* Electrons

# 46

Lucas K. Wagner

## Contents

### Abstract

This article briefly summarizes how to use standard fixed-node diffusion Monte Carlo to obtain accurate results for materials containing *d* and *f* electrons.

## 1  Introduction

Quantum Monte Carlo (QMC) techniques treat electron correlation directly and to reasonably high accuracy. Systems containing *d* and *f* electrons tend to have strong electron correlation effects, so one might expect that QMC techniques could be helpful to treating these systems. For the purposes of this article, we will mainly

L. K. Wagner (✉)
Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA
e-mail: lkwagner@illinois.edu

consider diffusion Monte Carlo (DMC), since most applications to $d$ and $f$ electron systems have used that technique. For clarity, we will also focus on standard techniques that have seen wide use. There are many variations on QMC methods that can sometimes offer higher accuracy but are less well-explored. For details on the methods, see the handbook article on QMC methods. The field is rather quickly moving, so this article will focus on basic considerations when performing calculations on these materials and some expectation for what can be achieved with standard approaches. It is not meant to be a comprehensive summary of the literature on this subject.

## 2 Technical Considerations for High Accuracy

Here we will list the current state of the art in performing these calculations to the highest accuracy and efficiency. Some of the guidelines here are rules of thumb that have been empirically found to work, while others are more rigorous. We will try to be as clear as possible on the difference between these two situations.

### 2.1 Effective Core Potentials/Pseudopotentials

Using effective core potentials based on density functional theory is particularly problematic for systems containing $d$ and $f$ electrons. Instead, high accuracy effective potentials so far seem to be based mainly on Hartree-Fock and quantum chemistry results. A reasonable justification for this is that the core electrons are described reasonably well in Hartree-Fock. This is because in the core, the electrons are at high density, in which case the terms covered exactly in Hartree-Fock dominate (see Ceperley and Alder 1980). Ultimately, one should show that the effective potentials reproduce high accuracy solutions of the all-electron atom and molecules to have confidence in the effective potentials. There are several repositories of high-quality effective core potentials designed for quantum Monte Carlo calculations available to date, which are based on the preceding principles. For some examples, see Burkatzki et al. (2007) and Trail and Needs (2015).

### 2.2 Effective Core Potentials/Pseudopotential Projection Error

In DMC, non-local potentials lead to an additional sign problem. The projection onto a low-energy state must then be performed approximately, leading to an additional dependence on the trial wave function beyond the fixed-node approximation. In recent years, this error has been investigated, and it can at times be a sizable fraction of the total error of the methods. In most cases, the projection error can be minimized by using high-quality Jastrow factors and energy optimization of the parameters.

## 2.3    Nodal Error/Orbitals

DMC requires a trial function to determine the nodal surface. Given a Hamiltonian, this nodal surface is one of the main approximations in the calculation. While it is possible to use complex trial wave functions to obtain high accuracy for small systems, for larger systems, typically a single Slater determinant is used as a trial wave function. For systems containing *d* and *f* electrons, the electron interactions are strong enough that they can affect the degree of hybridization in the one-particle orbitals. For that reason, orbitals based on Hartree-Fock methods or local density approximation are not typically optimal (see Wagner and Mitas 2007).

While direct orbital optimization is possible, it is often computationally prohibitive for large systems. It has been found that orbitals from hybrid density functionals are close to optimal (see Busemeyer et al. 2016) when compared to direct orbital optimization. In this technique, the amount of hybridization is varied, and a sequence of DMC calculations is performed. The orbitals that produce the lowest DMC energies give the lowest upper bound. Some authors also perform DFT+U calculations, but to the knowledge of this author, the relative performance of hybrid functionals versus a +U calculation has not been systematically assessed. Figure 1 contains the density difference between PBE0 (hybrid) and PBE orbitals for $La_2CuO_4$. A major theme is that the total charge on the Cu atom is larger in the PBE functional, which leads to worse nodes for the QMC calculation and significantly higher fixed-node energy.

## 2.4    Finite Size Effects

As with any many-body technique, we must extrapolate to infinite size to simulate a bulk material. There are a number of detailed papers on finite size scaling in DMC, for example Chiesa et al. (2006).



**Fig. 1** Density difference between a Slater determinant of hybrid PBE0 orbitals and PBE orbitals for $LaCu_2O_4$. The slice is through the copper oxide plane. Red regions indicate that PBE orbitals have higher density in the region, while blue means that PBE0 has higher density
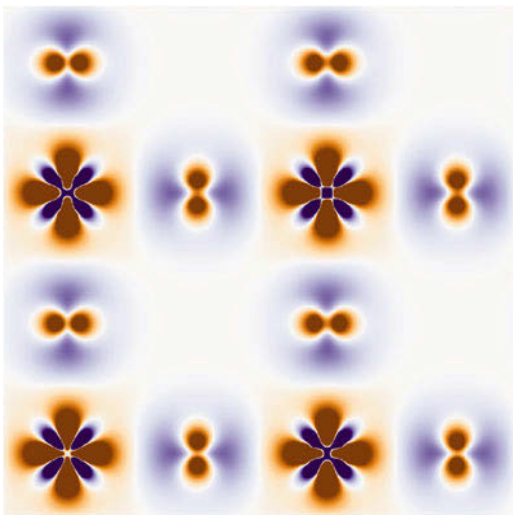
# 3    What Can Be Calculated Using DMC

- Ground state properties: magnetic/nonmagnetic ground state
- Spin properties: constrain to different spin states + model fitting

## 3.1    Ground State Energy as a Function of Atomic Position

FN-DMC offers very high accuracy on ground state properties, including those of $d$ and $f$ materials. As one can see in Fig. 2, it offers similar accuracy on those sorts of materials as for $s$ and $p$ systems, with errors of a few percent on many ground state properties.

## 3.2    Electronic Gap Calculation

One can consider two possible electronic gaps in interacting electronic systems. The first is called the *quasiparticle* or sometimes *transport* gap. This is computed using three ground state calculations of varying number of electrons:

$$\Delta_{QP} = E(N+1) + E(N-1) - 2E(N), \tag{1}$$

where $N$ is the number of electrons in a neutral unit cell and $E$ is the estimated ground state energy with that number. This quantity has the advantage that it involves only ground state calculations, so the variational principle obviously holds, but it can have large finite size effects.

The second gap is often called the *optical* gap. This is computed by promoting an electron from an unoccupied orbital to an occupied orbital. The fixed-node condition
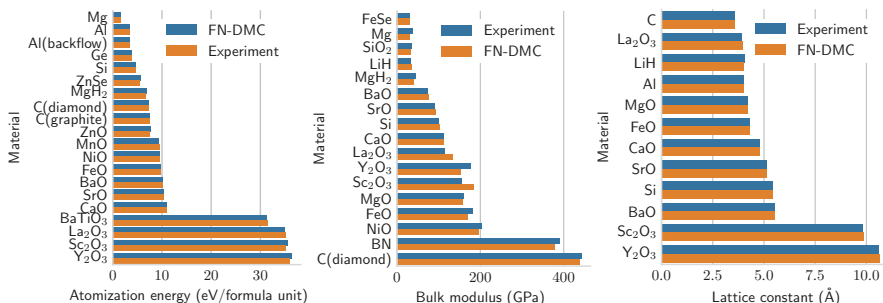


**Fig. 2** Performance of FN-DMC on several ground state properties for a variety of materials. (Data from Kolorenč and Mitas 2008, 2011, Santana et al. 2015, 2016, Mitra et al. 2015, Lee et al. 2004, Foyevtsova et al. 2014, Wagner and Abbamonte 2014, Zheng and Wagner 2015, Schiller et al. 2015, Yu et al. 2015, Hood et al. 2012, Pozzo and Alfè 2008, Shin et al. 2014, and Esler et al. 2010)

**Fig. 3** Electronic gap estimation for a variety of correlated electron materials
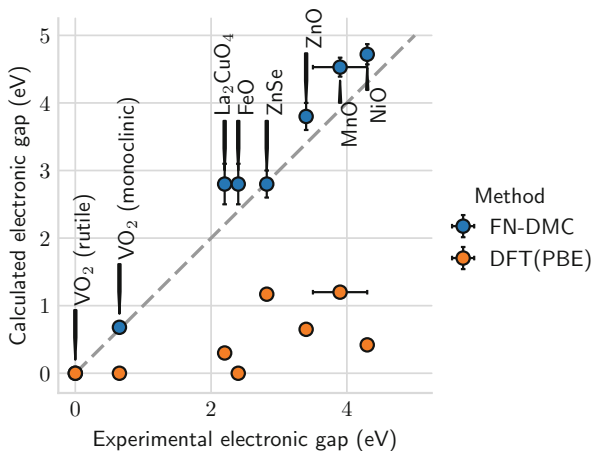
is used to prevent the system from falling into the ground state, so this can be viewed as an approximate computation of the excited state energy of the fixed $N$ system. This quantity approximates the lowest energy electronic optical transition that a material might have. It typically has smaller finite size effects and is easier to compute than the quasiparticle gap.

Figure 3 shows gaps computed for a number of correlated electron systems. Density functional theory calculations (DFT) in the most common PBE approximation are shown for reference. Most of these gaps were computed using the optical excitations, which match the transport gap typically to the accuracy of the calculation. The optical and the quasiparticle gap are not guaranteed to be the same value in general; however, they represent different physical quantities.

## 3.3    Nature of the Ground State

Since FN-DMC obtains tight upper bounds to the ground state energy, it can offer higher certainty estimates of the nature of the ground state than less accurate methods. To see why, suppose that one is interested in determining whether the ground state has double occupancy on a given atom in the ground state or not. If it has no double occupancy, then one might say that there are local spin moments and some sort of magnetic behavior would be expected, while in the contrary case, there are no local spin moments. Exactly this question has been asked about the metallic state of $VO_2$, which was addressed using fixed-node quantum Monte Carlo in Zheng and Wagner (2015).

To answer this question, consider a Hamiltonian $\hat{H}$. Divide all wave functions (which may or may not be eigenstates) into three classes:

$A$: double occupancy on the transition metal atom,
$B$: no double occupancy on the transition metal atom, and
$C$: superposition of $A$ and $B$.

**Fig. 4** Assignment of the
ground state class as a
function of the errors $\varepsilon$ in the
states. In this, $\Delta$ is the correct
difference between the actual
minima of the different
classes: $\Delta = E_{min}(B) -
E_{min}(A)$ for an $A$ ground
state
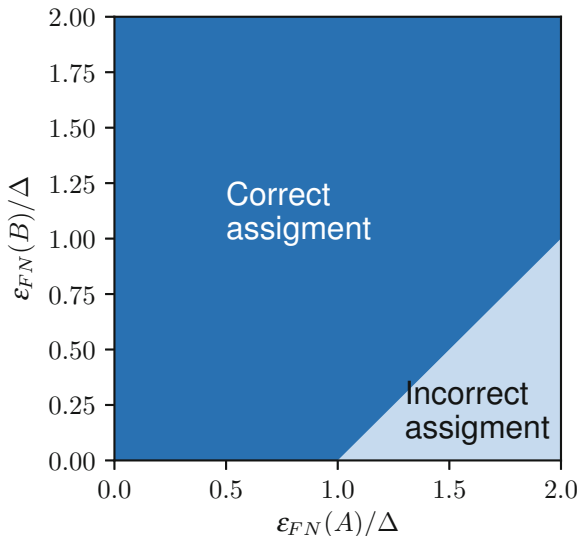


Define the function $E_{min}(X)$ as the minimum energy expectation value across all wave functions in class $X$.

In a fixed-node diffusion Monte Carlo calculation, one can typically control which class of wave function is produced using the trial function. Often several trial wave functions from each class are tested; let the minimum fixed-node energy for each class be $E_{FN}(A)$, $E_{FN}(B)$. One should be careful to evaluate the class after the diffusion Monte Carlo process. If the nodal optimization is perfect, then $E_{FN}(X) = E_{min}(X)$ for class $X$; however, in general $E_{FN}(X) \geq E_{min}(X)$. The error of a class $X$ is $\varepsilon_{FN}(X) = E_{FN}(X) - E_{min}(X)$.

Assume without loss of generality that $E_{min}(A) < E_{min}(C) < E_{min}B$. In that case, the ground state is of class $A$. To decide between $A$ and $B$, one evaluates

$$E_{FN}(B) - E_{FN}(A) = E_{min}(B) - E_{min}(A) + \varepsilon_{FN}(A) - \varepsilon_{FN}(B). \tag{2}$$

As one can easily see that the smaller the $\varepsilon$ values (more accurate calculations), the less likely a misidentification of the ground state (Fig. 4).

## 4      Summary

Fixed-node quantum Monte Carlo calculations offer highly accurate results for materials containing $d$ and $f$ elements. Since they treat the electron correlation explicitly and accurately, they can predict many properties of materials, some of which are summarized in this article.

# References

Burkatzki M, Filippi C, Dolg M (2007) Energy-consistent pseudopotentials for quantum Monte Carlo calculations. J Chem Phys 126(23):234105–234105–8. https://doi.org/doi:10.1063/1.2741534, http://jcp.aip.org/resource/1/jcpsa6/v126/i23/p234105_s1

Busemeyer B, Dagrada M, Sorella S, Casula M, Wagner LK (2016) Competing collinear magnetic structures in superconducting FeSe by first principles quantum Monte Carlo calculations. Phys. Rev. B 94:035108

Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45(7):566–569. http://link.aps.org/doi/10.1103/PhysRevLett.45.566

Chiesa S, Ceperley D, Martin R, Holzmann M (2006) Finite-size error in many-body simulations with long-range interactions. Phys Rev Lett 97(7). http://link.aps.org/doi/10.1103/PhysRevLett.97.076404

Esler KP, Cohen RE, Militzer B, Kim J, Needs RJ, Towler MD (2010) Fundamental high-pressure calibration from all-electron quantum Monte Carlo calculations. Phys Rev Lett 104(18):185702. http://link.aps.org/doi/10.1103/PhysRevLett.104.185702

Foyevtsova K, Krogel JT, Kim J, Kent PRC, Dagotto E, Reboredo FA (2014) Ab initio quantum Monte Carlo calculations of spin superexchange in cuprates: the Benchmarking case of $Ca_2CuO_3$. Phys Rev X 4(3):031003. http://link.aps.org/doi/10.1103/PhysRevX.4.031003

Hood RQ, Kent PRC, Reboredo FA (2012) Diffusion quantum Monte Carlo study of the equation of state and point defects in aluminum. Phys Rev B 85(13):134109. http://link.aps.org/doi/10.1103/PhysRevB.85.134109

Kolorenč J, Mitas L (2008) Quantum Monte Carlo calculations of structural properties of FeO under pressure. Phys Rev Lett 101(18):185502. http://link.aps.org/doi/10.1103/PhysRevLett.101.185502

Kolorenč J, Mitas L (2011) Applications of quantum Monte Carlo methods in condensed systems. Rep Prog Phys 74(2):026502. https://doi.org/10.1088/0034-4885/74/2/026502, http://iopscience.iop.org/0034-4885/74/2/026502

Lee JW, Mitas L, Wagner LK (2004) Quantum Monte Carlo study of MnO solid. arXiv:cond-mat/0411247. http://arxiv.org/abs/cond-mat/0411247

Mitra C, Krogel JT, Santana JA, Reboredo FA (2015) Many-body ab initio diffusion quantum Monte Carlo applied to the strongly correlated oxide NiO. J Chem Phys 143(16):164710. https://doi.org/10.1063/1.4934262, http://scitation.aip.org/content/aip/journal/jcp/143/16/10.1063/1.4934262

Pozzo M, Alfè D (2008) Structural properties and enthalpy of formation of magnesium hydride from quantum Monte Carlo calculations. Phys Rev B 77(10):104103. http://link.aps.org/doi/10.1103/PhysRevB.77.104103

Santana JA, Krogel JT, Kim J, Kent PRC, Reboredo FA (2015) Structural stability and defect energetics of ZnO from diffusion quantum Monte Carlo. J Chem Phys 142(16):164705. https://doi.org/10.1063/1.4919242, http://scitation.aip.org/content/aip/journal/jcp/142/16/10.1063/1.4919242

Santana JA, Krogel JT, Kent PRC, Reboredo FA (2016) Cohesive energy and structural parameters of binary oxides of groups IIA and IIIB from diffusion quantum Monte Carlo. J Chem Phys 144(17):174707. https://doi.org/10.1063/1.4947569, http://scitation.aip.org/content/aip/journal/jcp/144/17/10.1063/1.4947569

Schiller JA, Wagner LK, Ertekin E (2015) Phase stability and properties of manganese oxide polymorphs: assessment and insights from diffusion Monte Carlo. Phys Rev B 92(23):235209. http://link.aps.org/doi/10.1103/PhysRevB.92.235209

Shin H, Kang S, Koo J, Lee H, Kim J, Kwon Y (2014) Cohesion energetics of carbon allotropes: quantum Monte Carlo study. J Chem Phys 140(11):114702. https://doi.org/10.1063/1.4867544, http://scitation.aip.org/content/aip/journal/jcp/140/11/10.1063/1.4867544

Trail JR, Needs RJ (2015) Correlated electron pseudopotentials for 3D-transition metals. J Chem Phys 142(6):064110. https://doi.org/10.1063/1.4907589, http://aip.scitation.org/doi/abs/10.1063/1.4907589

Wagner LK, Abbamonte P (2014) Effect of electron correlation on the electronic structure and spin-lattice coupling of high-Tc cuprates: quantum Monte Carlo calculations. Phys Rev B 90(12):125129. http://link.aps.org/doi/10.1103/PhysRevB.90.125129

Wagner LK, Mitas L (2007) Energetics and dipole moment of transition metal monoxides by quantum Monte Carlo. J Chem Phys 126(3):034105–034105–5. https://doi.org/doi:10.1063/1.2428294, http://jcp.aip.org.proxy2.library.illinois.edu/resource/1/jcpsa6/v126/i3/p034105_s1

Yu J, Wagner LK, Ertekin E (2015) Towards a systematic assessment of errors in diffusion Monte Carlo calculations of semiconductors: case study of zinc selenide and zinc oxide. J Chem Phys 143(22):224707. https://doi.org/10.1063/1.4937421, http://scitation.aip.org/content/aip/journal/jcp/143/22/10.1063/1.4937421

Zheng H, Wagner LK (2015) Computation of the Correlated metal-insulator transition in vanadium dioxide from first principles. Phys Rev Lett 114(17):176401. http://link.aps.org/doi/10.1103/PhysRevLett.114.176401

Pui-Wai Ma  and S. L. Dudarev

## Contents

**Abstract**

Finite temperature magnetic fluctuations determine a variety of properties of magnetic materials, including their phase stability, their thermodynamic properties, and even the structure of defects formed under irradiation. A fundamental feature of microscopic magnetic fluctuations is the directional non-collinearity of fluctuating atomic magnetic moments, which stems from the rotational invariance of an atomic magnetic Hamiltonian. To model the dynamics of magnetic moments of atoms that move themselves, a fast and computationally efficient simulation approach is required. Spin-lattice dynamics simulates atomic movements as well as rotational and longitudinal fluctuations of atomic magnetic moments within a unified framework, generalizing molecular dynamics

P.-W. Ma (✉) · S. L. Dudarev
Culham Centre for Fusion Energy, UK Atomic Energy Authority, Culham Science Centre, Abingdon, UK
e-mail: Leo.Ma@ukaea.uk

to magnetic materials. Collective magnetic and atomic excitations can now be investigated on the microscopic scale, similarly to how transformations of atomic structures can be investigated using molecular dynamics simulations. This chapter outlines theoretical foundations and numerical algorithms of spin-lattice dynamics and describes applications of the method.

# 1    Introduction

Magnetic materials play a pivotal part in modern technology; their applications include long-term information storage, fast access memory devices, and even quantum computing. Magnetic and mechanical properties of magnetic materials are intimately related, in particular this applies to steels and iron alloys. For example, ferromagnetic iron is the only element in Group 8 of the Periodic Table that adopts the bcc crystal structure. All the other elements in the same group have hcp crystal structure, and bcc Fe owes its stability to magnetism (Pettifor 1995). A self-interstitial atom defect in bcc iron adopts a ⟨110⟩ dumbbell configuration (Fu et al. 2004), whereas in all the nonmagnetic bcc metals, a single self-interstitial defect has the ⟨111⟩ symmetry (Nguyen-Manh et al. 2006; Derlet et al. 2007). At high temperatures, bcc-fcc-bcc phase transitions in iron occur as a result of competition between magnetic excitations and atomic vibrations (Lavrentiev et al. 2010; Ma et al. 2017).

Neither molecular dynamics nor spin dynamics on their own can capture *both* magnetic and atomic excitations. A broader mathematical simulation framework is required to describe the dynamics of spin and lattice subsystems and their coupling. Omelyan et al. (2001a, b, 2002) and Tsai et al. (2004, 2005) proposed models unifying spin dynamics and molecular dynamics. We have developed their ideas further, arriving at an algorithm suitable for simulating real materials.

Spin-lattice dynamics follows the time evolution of coupled spin and lattice subsystems. Precession of spins and atomic motion are coupled through spin-dependent forces and coordinate-dependent effective exchange fields. Figure 1 shows snapshots of magnetic order in iron at 0 and 300 K. The figure illustrates an important aspect of finite temperature magnetic simulations. Magnetic moments (which sometimes are also called atomic spins) become non-collinear due to thermal excitation. The fact that interaction between magnetic moments depends on the position of atoms, generates additional, spin direction dependent, forces acting between the atoms in the material. This is how magnetic excitations affect the dynamics and stability of atomic lattice.

In what follows, we first discuss the spin equations of motion and the notion of spin temperature. Then, we outline theoretical foundations of spin-lattice dynamics. We also describe a numerical integration algorithm that does not normally receive much attention in the context of molecular dynamics but proves essential in the framework of spin-lattice dynamics. Finally, we highlight the still outstanding scientific challenges, particularly those associated with magnetic many-body interatomic interaction potentials.
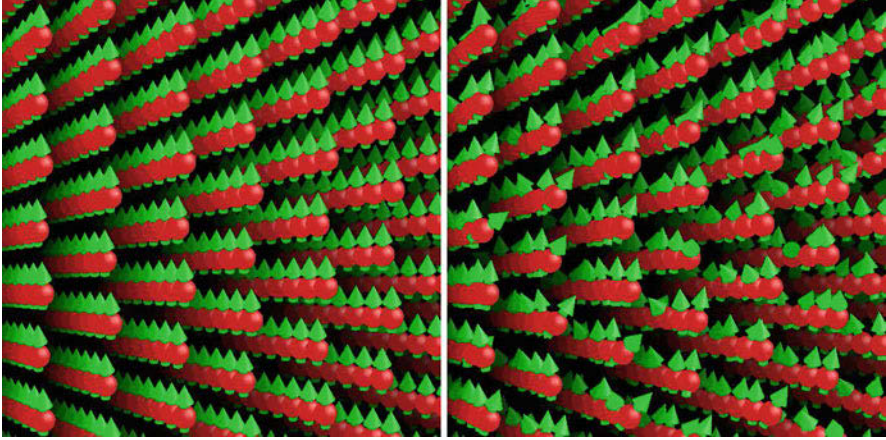
**Fig. 1** Snapshots of magnetic configurations generated using spin-lattice dynamics simulations of iron at 0 and 300 K. Simulations were performed using spin-lattice dynamics simulation program SPILADY (Ma et al. 2016)

## 2     The Quantum Origin of Spin Dynamics

The localized nature of $d$ and $f$ electrons justifies the notion of an atomic magnetic moment **M** or atomic spin **S**, where $\mathbf{M} = -g\mu_B\mathbf{S}$, $g$ is the electron $g$-factor, and $\mu_B$ is the Bohr magneton. Atomic magnetic moments form due to intra-atomic exchange interaction between the localized electrons. Interaction between magnetic moments associated with different atoms can be understood as resulting from the interplay between intra-atomic exchange and interatomic quantum hopping of electrons. The quantum nature of atomic spins gives rise to their unusual properties, for example, an atomic spin has no mass and hence has no conjugate variable. The classical equation of motion for a spin vector is in fact the mean-field analog of the quantum equation of motion for the spin operator, which can be derived using the Poisson brackets approach (Ma and Dudarev 2012).

Consider an arbitrary quantum-mechanical spin Hamiltonian $\hat{\mathscr{H}}$. It can be expressed as a Taylor series in the spin vector operator $\hat{\mathbf{S}}$ as

$$\hat{\mathscr{H}} = \sum_{n=0}^{\infty} \mathbf{a}_n \hat{\mathbf{S}}^n, \tag{1}$$

where tensor quantities $\mathbf{a}_n$ are the Taylor series expansion coefficients. The $n^{th}$ order of operator $\hat{\mathbf{S}}$ can be written as

$$\begin{aligned}
\hat{\mathbf{S}}^n &= (\hat{\mathbf{I}}\mathbf{S} + \delta\hat{\mathbf{S}})^n, \\
&= \hat{\mathbf{I}}\mathbf{S}^n + n\mathbf{S}^{n-1}\delta\hat{\mathbf{S}} + \cdots,
\end{aligned} \tag{2}$$

where $\mathbf{S}$ is the expectation value of $\hat{\mathbf{S}}$, $\hat{\mathbf{I}}$ is the identity operator, and $\delta\hat{\mathbf{S}} = \hat{\mathbf{S}} - \hat{\mathbf{I}}\mathbf{S}$ is what can be called the deviation of the operator from its expectation value. Substituting Eq. (2) into Eq. (1), we find

$$\hat{\mathscr{H}} = \hat{\mathbf{I}} \sum_{n=0}^{\infty} \mathbf{a}_n \mathbf{S}^n + \sum_{n=1}^{\infty} \mathbf{a}_n n \mathbf{S}^{n-1} \delta\hat{\mathbf{S}} + \cdots ,$$

$$= \hat{\mathbf{I}}\mathscr{H} + \frac{\partial\mathscr{H}}{\partial\mathbf{S}} \cdot \delta\hat{\mathbf{S}} + \cdots . \tag{3}$$

The first term in (3) is the Hamiltonian *function* $\mathscr{H}$, which is equivalent to a classical Hamiltonian, where the spin operator $\hat{\mathbf{S}}$ is replaced by its expectation value $\mathbf{S}$. Ignoring higher-order terms in Eq. (3), the equation of motion for a spin operator can be derived using the Poisson brackets commutator

$$\frac{d\hat{\mathbf{S}}}{dt} = \frac{i}{\hbar}\left[\hat{\mathscr{H}}, \hat{\mathbf{S}}\right] = \frac{1}{\hbar}\left[\hat{\mathbf{S}} \times \left(-\frac{\partial\mathscr{H}}{\partial\mathbf{S}}\right)\right],$$

$$= \frac{1}{\hbar}\left[\hat{\mathbf{S}} \times \mathbf{H}\right], \tag{4}$$

where $\mathbf{H} = -\partial\mathscr{H}/\partial\mathbf{S}$ is the effective field acting on spin $\hat{\mathbf{S}}$. Since the first term in Eq. (3) commutes with $\hat{\mathbf{S}}$, it gives no contribution to the equation of motion. The first non-vanishing contribution to the right-hand side of (4) comes from the second term in Eq. (3).

The above derivation remains valid for any spin Hamiltonian. We note that the form of Eq. (4) is the same as that of the equation of motion for a classical spin vector. This can be proven by evaluating expectation values of both sides of the equation.

## 3    Spin Temperature Monitoring and Control

If a statistical and dynamically evolving system is in contact with another statistical, and also dynamically evolving, system, energy flows from the hotter to the cooler one, until they reach thermal equilibrium. There is an established procedure for thermalizing an atomic system in a molecular dynamics simulation, which involves putting the system in contact with a heat reservoir, represented by certain fluctuation and dissipation terms in the classical equations of motion for the atoms. This thermalization method is known as Langevin dynamics. The treatment of Brownian motion is probably one of the best known examples of application of the method (Chandrasekhar 1943; Kubo 1966). The use of Langevin dynamics for equilibrating and thermalizing large systems of interacting atoms is a well-established part of the molecular dynamics simulation toolkit.

In the preceding section, we derived equations of motion for a closed system of interacting spins. To thermalize a spin system, one can also use a special form of Langevin dynamics, the implementation of which requires suitably modified spin equations of motion. Langevin dynamics approach to spin dynamics was proposed by Brown (1963). For an arbitrary Hamiltonian $\mathscr{H}(\{\mathbf{S}_i\})$ describing $N$ atomic spins, the Langevin equation of motion for an atomic spin vector $i$ can be written as

$$\frac{d\mathbf{S}_i}{dt} = \frac{1}{\hbar}\left[\mathbf{S}_i \times (\mathbf{H}_i + \mathbf{h}_i) - \gamma\mathbf{S}_i \times (\mathbf{S}_i \times \mathbf{H}_i)\right] \tag{5}$$

where $\mathbf{H}_i = -\partial\mathscr{H}/\partial\mathbf{S}_i$ is the effective magnetic field acting on atomic spin $i$, $\gamma$ is a dissipation constant, and $\mathbf{h}_i$ is a delta-correlated fluctuating vector satisfying conditions $\langle\mathbf{h}_i(t)\rangle = 0$ and $\langle h_{i\alpha}(t)h_{j\beta}(t')\rangle = \mu\delta_{ij}\delta_{\alpha\beta}\delta(t - t')$. Subscripts $\alpha$ and $\beta$ refer to the Cartesian components of a vector, and parameter $\mu$ characterizes the magnitude of thermal magnetic fluctuations.

Stochastic and dissipative forces acting together drive a dynamic system to thermal equilibrium. Fluctuation and dissipation terms are related through the fluctuation-dissipation theorem (Chandrasekhar 1943; Kubo 1966). The fluctuation-dissipation relation (FDR) between the fluctuating and dissipative terms can be obtained by mapping the Langevin equation of motion to a Fokker-Planck equation (Zwanzig 2001; Van Kampen 2011) and identifying the asymptotic stationary solution of that equation with the Gibbs distribution (Brown 1963; Ma and Dudarev 2011). The FDR for Eq. (5) reads $\mu = 2\gamma\hbar k_B T$, where $T$ is the temperature of the heat reservoir. Without the fluctuating term, Eq. (5) reduces to the Landau-Lifshitz equation (Landau and Lifshitz 1935; Gilbert 2004), which contains only the dissipative term. In the asymptotic limit $t \to \infty$, solutions of the Landau-Lifshitz equation describe stationary spin configurations, where $d\mathbf{S}_i/dt = 0$. Directions of vectors $\mathbf{S}_i$ in a stationary spin configuration can be found by solving equations $\mathbf{S}_i \times \mathbf{H}_i = 0$, where $i = 1, 2 \ldots N$ (see Lavrentiev et al. 2011). The latter condition has a simple meaning, namely, that in the lowest energy configuration every spin vector $\mathbf{S}_i$ is collinear with the exchange field $\mathbf{H}_i$ acting on it.

A notable feature of Eq. (5) is that the magnitude of the spin vector $|\mathbf{S}_i(t)|$, where $\mathbf{S}_i(t)$ is a solution of the equation, remains constant. This can be easily proven by multiplying both sides of the equation by $\mathbf{S}_i$ and noting that a vector product, involving an arbitrary vector, is orthogonal to it. As a result, the magnitude of the spin vector is conserved $d\mathbf{S}_i^2(t)/dt = 0$.

Due to the simultaneously localized and itinerant nature of electrons in a solid, both the magnitude and direction of atomic spins are variable quantities. A revision of Langevin spin dynamics is required to relax the constraint that the magnitude of an evolving spin vector is a constant. Bearing in mind the Langevin treatment of atomic dynamics, we find that longitudinal fluctuations of an atomic magnetic moment, i.e., fluctuations of the magnitude of a spin vector, can indeed be treated using some suitably chosen fluctuation and dissipation terms (Ma and Dudarev 2012). We write

$$\frac{d\mathbf{S}_i}{dt} = \frac{1}{\hbar}\left[\mathbf{S}_i \times \left(-\frac{\partial \mathscr{H}}{\partial \mathbf{S}_i}\right)\right] - \gamma\frac{\partial \mathscr{H}}{\partial \mathbf{S}_i} + \boldsymbol{\xi}_i, \tag{6}$$

$$= \frac{1}{\hbar}\left[\mathbf{S}_i \times \mathbf{H}_i\right] + \gamma\mathbf{H}_i + \boldsymbol{\xi}_i, \tag{7}$$

where $\boldsymbol{\xi}_i$ is a delta-correlated fluctuating vector satisfying conditions $\langle\boldsymbol{\xi}_i(t)\rangle = 0$ and $\langle\xi_{i\alpha}(t)\xi_{j\beta}(t')\rangle = \mu\delta_{ij}\delta_{\alpha\beta}\delta(t-t')$. The FDR can now be obtained by mapping Eq. (6) onto a Fokker-Planck equation (Zwanzig 2001; Van Kampen 2011):

$$\frac{\partial W}{\partial t} = -\sum_{i\alpha}\frac{\partial}{\partial S_{i\alpha}}(A_{i\alpha}W) + \frac{1}{2}\sum_{ij\alpha\beta}\frac{\partial^2}{\partial S_{i\alpha}\partial S_{j\beta}}(B_{i\alpha j\beta}W). \tag{8}$$

In the above equation, $A_{i\alpha} = \lim_{\Delta t \to 0}\frac{1}{\Delta t}\langle S_{i\alpha}\rangle$ is an effective drift coefficient, and $B_{i\alpha j\beta} = \lim_{\Delta t \to 0}\frac{1}{\Delta t}\langle S_{i\alpha}S_{j\beta}\rangle$ is an effective diffusion coefficient. According to Eq. (6), the drift and diffusion coefficients have the form (Zwanzig 2001; Van Kampen 2011):

$$\mathbf{A}_i = \frac{1}{\hbar}\left[\mathbf{S}_i \times \left(-\frac{\partial \mathscr{H}}{\partial \mathbf{S}_i}\right)\right] - \gamma\frac{\partial \mathscr{H}}{\partial \mathbf{S}_i}, \tag{9}$$

$$B_{i\alpha j\beta} = \mu\delta_{ij}\delta_{\alpha\beta}. \tag{10}$$

In thermal equilibrium, the energy distribution asymptotically approaches the Gibbs distribution $W = W_0\exp(-\mathscr{H}/k_BT)$, where $W_0$ is a normalization constant. Substituting Eqs. (9) and (10), and the Gibbs distribution, into Eq. (8), one finds that

$$\frac{\partial W}{\partial t} = \left(\gamma - \frac{\mu}{2k_BT}\right)\left[\sum_{i,\alpha}\left(\frac{\partial^2 \mathscr{H}}{\partial S_{i\alpha}^2} - \frac{1}{k_BT}\left(\frac{\partial \mathscr{H}}{\partial S_{i\alpha}}\right)^2\right)\right]W. \tag{11}$$

Stationary solutions of this equation corresponding to $\partial W/\partial t = 0$ describe thermal equilibrium. From Eq. (11) we see that the right-hand side of Eq. (11) vanishes if

$$\mu = 2\gamma k_BT. \tag{12}$$

Surprisingly, the form of this FDR is exactly the same as that of the lattice Langevin dynamics. The right-hand side of Eq. (11) also vanishes if

$$k_BT = \sum_{i,\alpha}\left(\frac{\partial \mathscr{H}}{\partial S_{i\alpha}}\right)^2 \bigg/ \sum_{i,\alpha}\frac{\partial^2 \mathscr{H}}{\partial S_{i\alpha}^2}. \tag{13}$$

Equation (13) defines the dynamic spin temperature at equilibrium as a function of microscopic dynamic variables. This resembles the well-known equipartition principle for atoms where the lattice temperature can be estimated using the relation $3Nk_BT/2 = \sum_i\mathbf{P}_i^2/2m$, where $\mathbf{P}_i$ is the momentum of atom $i$. Equation (13)

has similar functionality, and it can also be used for estimating the local spin temperature of an arbitrary spin configuration, which potentially may be very far from equilibrium. Applying a similar procedure to Eq. (5), we find an alternative formula for the spin temperature, which applies if the longitudinal fluctuations of magnetic moments can be neglected (Ma et al. 2010).

$$k_B T = \frac{\sum_i |\mathbf{S}_i \times \mathbf{H}_i|^2}{2 \sum_i \mathbf{S}_i \cdot \mathbf{H}_i}. \tag{14}$$

Here we note an alternative way of deriving Eqs. (13) and (14). If a system is in thermal equilibrium, there is no net energy exchange with the heat reservoir, resulting in $d\langle E \rangle/dt = 0$. Since $dE/dt = \sum_i (\partial E/\partial \mathbf{S}_i)(d\mathbf{S}_i/dt)$, by using either of the two forms of Langevin equations of motion for the spins, taking the ensemble average, and applying the FDR, we arrive at the above formulae for the spin temperature expressed in terms of dynamic spin variables. Condition $T = 0$ corresponds to the lowest energy spin configuration that, according to Eq. (14), is defined by a set of algebraic equations noted earlier in this section, namely, $\mathbf{S}_i \times \mathbf{H}_i = 0$ for $i = 1, 2 \ldots, N$.

## 4    Spin-Lattice Dynamics

Interaction between atoms in a magnetic material is determined by its spin-dependent electronic structure. A suitable mathematical framework is required to describe the many-body phonon and magnon excitations at elevated temperature. Conventional molecular dynamics provides a convenient starting point for the incorporation of spin degrees of freedom in an atomistic simulation. It is possible to reformulate molecular dynamics and spin dynamics and combine them within a unified simulation framework. This also makes it possible to treat interaction between the lattice and magnetic subsystems, where interatomic forces and effective exchange fields acting on magnetic moments are related and dynamically coupled with each other.

Consider an arbitrary coordinate and spin-dependent Hamiltonian $\mathcal{H}(\mathbf{R}, \mathbf{P}, \mathbf{S})$, where $\mathbf{R} = \{\mathbf{R}_i\}$ are the atomic coordinates, $\mathbf{P} = \{\mathbf{P}_i\}$ are the atomic momenta, and $\mathbf{S} = \{\mathbf{S}_i\}$ are the atomic spin vectors. The Hamilton equations of motion for a closed system have the form (Ma et al. 2008, 2016):

$$\begin{aligned}
\frac{d\mathbf{R}_i}{dt} &= \frac{\partial \mathcal{H}}{\partial \mathbf{P}_i}, \\
\frac{d\mathbf{P}_i}{dt} &= -\frac{\partial \mathcal{H}}{\partial \mathbf{R}_i}, \\
\frac{d\mathbf{S}_i}{dt} &= \frac{1}{\hbar} \left[ \mathbf{S}_i \times \left( -\frac{\partial \mathcal{H}}{\partial \mathbf{S}_i} \right) \right].
\end{aligned} \tag{15}$$

From their appearance, these equations formally amount to no more than a combination of molecular dynamics and spin dynamics, and the only quantity that formally unifies them is the Hamiltonian. However, there is a subtle difference between spin-lattice dynamics defined by Eqs. (15) above and molecular dynamics or spin dynamics treated as separate simulation methods. The force $\mathbf{F}_i = -\partial \mathscr{H}/\partial \mathbf{R}_i$ acting on atom $i$ and the effective magnetic field $\mathbf{H}_i = -\partial \mathscr{H}/\partial \mathbf{S}_i$ acting on spin vector $i$ are coordinate *and* spin-dependent. The direction and magnitude of atomic spin vectors $\mathbf{S}_i$ affect the direction and magnitude of forces acting between the atoms in a magnetic material, whereas the effective exchange field acting on spin $\mathbf{S}_i$ depends on atomic positions. The lattice and spin subsystems are now coupled through the spin-orientation-dependent interatomic forces and coordinate-dependent effective exchange fields.

Generalizing Eqs. (15) to Langevin dynamics, we write the Langevin equations of motion for the spins and atomic coordinates and momenta as

$$\frac{d\mathbf{R}_i}{dt} = \frac{\partial \mathscr{H}}{\partial \mathbf{P}_i},$$

$$\frac{d\mathbf{P}_i}{dt} = -\frac{\partial \mathscr{H}}{\partial \mathbf{R}_i} - \gamma_l \frac{\partial \mathscr{H}}{\partial \mathbf{P}_i} + \mathbf{f}_i,$$

$$\frac{d\mathbf{S}_i}{dt} = \frac{1}{\hbar} \left[ \mathbf{S}_i \times \left( -\frac{\partial \mathscr{H}}{\partial \mathbf{S}_i} \right) \right] - \gamma_s \frac{\partial \mathscr{H}}{\partial \mathbf{S}_i} + \boldsymbol{\xi}_i, \tag{16}$$

where $\mathbf{f}_i$ and $\boldsymbol{\xi}_i$ are the fluctuating components of interatomic forces and exchange fields, respectively. They satisfy the Langevin equations conditions $\langle \mathbf{f}_i(t) \rangle = 0$ and $\langle f_{i\alpha}(t) f_{j\beta}(t') \rangle = \mu_l \delta_{ij} \delta_{\alpha\beta} \delta(t - t')$ for the fluctuating components of interatomic forces and $\langle \boldsymbol{\xi}_i(t) \rangle = 0$ and $\langle \xi_{i\alpha}(t) \xi_{j\beta}(t') \rangle = \mu_s \delta_{ij} \delta_{\alpha\beta} \delta(t - t')$ for the fluctuating exchange fields. The FDR relations for the lattice and spin subsystems read $\mu_l = 2\gamma_l k_B T$ and $\mu_s = 2\gamma_s k_B T$, where $T$ is the temperature of the heat reservoir. The dissipative constants $\gamma_l$ and $\gamma_s$ determine the thermalization rates, which can be derived from experimental observations. An example of thermalization process is given in Fig. 2. The figure shows how the spin and lattice temperatures vary in an interacting spin-lattice dynamic system. The simulation involves 16,000 magnetic atoms of iron. The reservoir temperature is set to 300 K. We used the values of parameters describing ferromagnetic iron that were derived by Ma and Dudarev (2012) and Ma et al. (2012). The different thermalization rates characterizing spin and lattice subsystems are primarily due to the difference between the values of dissipation constants $\gamma_l$ and $\gamma_s$.

The spin-lattice dynamics simulation model can be extended to include the treatment of conduction electrons, if we assume that there is a significant amount of heat dissipated to the electrons. This is often the case in metals. For example, in applications, the heat reservoir is nothing but the time-dependent evolving electron subsystem. Its dynamic behavior is described by the heat transfer equation

$$C_e \frac{dT_e}{dt} = \nabla(\kappa_e \nabla T_e) + G_{el}(T_l - T_e) + G_{es}(T_s - T_e), \tag{17}$$

**Fig. 2** Spin and lattice temperatures in magnetic iron during thermalization, predicted by spin-lattice dynamics simulations. The temperature of the reservoir is 300 K. Initial temperatures of atoms and spins are $T = 0$ K. The simulation follows the evolution of 16,000 magnetic iron atoms, initially forming a perfect bcc lattice configuration
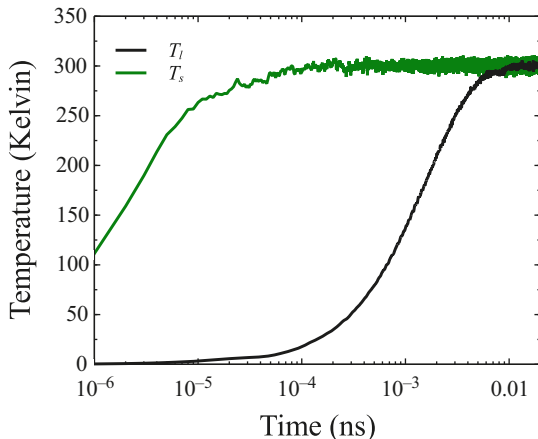
where $C_e$ is the electronic specific heat and $\kappa_e$ is the coefficient of thermal conductivity. The coupling constants describing interaction between electron and lattice subsystems $G_{el}$ and between electron and spin subsystems $G_{es}$ are

$$G_{el} = \frac{3k_B \gamma_l}{m\Omega} \tag{18}$$

$$G_{es} = \frac{k_B \gamma_s}{\Omega} \langle \sum_\alpha \frac{\partial^2 \mathcal{H}}{\partial S_{i\alpha}^2} \rangle \tag{19}$$

where $\Omega$ is the atomic volume. Detailed derivations of $G_{el}$ and $G_{es}$ are given in Ma et al. (2012, 2016) and references therein. Equations (16) and (17) constitute a fully self-consistent atomistic spin-lattice-electron model. In the ultrafast laser experiments, where over a relatively short period of time only a small amount of heat is exchanged with the environment, the spin-lattice-electron model can be successfully applied to modelling transient processes of thermalization and heat exchange between the three subsystems, as well as to the treatment of fast demagnetization (Ma et al. 2016).

## 5 The Suzuki-Trotter Decomposition

In a dynamic simulation of the evolution of a system of atoms, a numerical algorithm advances the state of the system as a function of time by propagating a configuration through a sequence of finite time steps. This generates numerical errors, for example, biased errors and truncation errors, resulting from analytical or numerical approximations. An integration algorithm based on the Suzuki-Trotter decomposition (STD) (Hatano and Suzuki 2005) can minimize numerical errors over a relatively long interval of computation time due to its symplectic nature,

which means that the algorithm conserves the phase space volume of the system during evolution. The STD involves breaking up an evolution operator, which consists of several noncommutative operations, into simpler sub-evolution steps. The second-order STD has the form

$$e^{(\hat{A}+\hat{B})\Delta t} = e^{\hat{A}\Delta t/2}e^{\hat{B}\Delta t}e^{\hat{A}\Delta t/2} + O(\Delta t^3) \tag{20}$$

where $\hat{A}$ and $\hat{B}$ are arbitrary operators and $\Delta t$ is the time step. The above formula describes how to decompose an operator that evolves the system over a time step, into several simpler evolution steps, each involving the evolution of only a subset of variables describing the microscopic configuration of the system.

Omelyan et al. (2001a, b, 2002) and Tsai et al. (2004, 2005) explored applications of the STD to spin-lattice dynamics, and in what follows we adopt a similar approach. Equations of motion for a spin-lattice system can be rewritten as

$$\frac{d\mathbf{x}}{dt} = (\mathcal{R} + \mathcal{P} + \mathcal{S})\mathbf{x}, \tag{21}$$

where $\mathbf{x} = \{\mathbf{R}, \mathbf{P}, \mathbf{S}\}$ is a generalized coordinate and $\mathcal{R}$, $\mathcal{P}$, $\mathcal{S}$ are the evolution operators acting on $\mathbf{R}$, $\mathbf{P}$, and $\mathbf{S}$, respectively. The formal solution of Eq. (21) can be written as

$$\mathbf{x}(t + \Delta t) = e^{(\mathcal{R}+\mathcal{P}+\mathcal{S})\Delta t}\mathbf{x}(t). \tag{22}$$

Using the STD decomposition given by Eq. (20), we write

$$e^{(\mathcal{R}+\mathcal{P}+\mathcal{S})\Delta t} = e^{\mathcal{P}\Delta t/2}e^{\mathcal{S}\Delta t/2}e^{\mathcal{R}\Delta t}e^{\mathcal{S}\Delta t/2}e^{\mathcal{P}\Delta t/2} + O(\Delta t^3). \tag{23}$$

Reading the right-hand side of this equation from right to left, we see that the STD decomposition rule requires that we would integrate equations for the momenta of particles over the time interval $\Delta t/2$, then integrate equations for the spins over the time interval $\Delta t/2$, and then integrate equations for atomic coordinates over the time interval $\Delta t$, followed by the integration of equations for the spins over $\Delta t/2$ and equations for the momenta over $\Delta t/2$. The order in which we integrate the equations minimizes the number of times where forces are evaluated and significantly reduces the time required to do a simulation. The main advantage offered by the STD (23) is that it circumvents the need to integrate the coupled equations for the coordinates, momenta, and spins all at the same time.

A particular subtlety associated with the presence of spin equations of motion in Eqs. (15) is that spin dynamics involve rotations, which, as opposed to translations, do not commute. Bearing this in mind, the evolution of the spin subsystem can be split into a series of operations involving evolution of individual spins, namely

$$e^{\mathcal{S}\Delta t} = e^{\mathcal{S}_1\Delta t/2}e^{\mathcal{S}_2\Delta t/2}\cdots e^{\mathcal{S}_N\Delta t}\cdots e^{\mathcal{S}_2\Delta t/2}e^{\mathcal{S}_1\Delta t/2} + O(\Delta t^3). \tag{24}$$

Each operation of evolution of an individual spin depends on all the previous operations, because the effective magnetic field $\mathbf{H}_i$ acting on spin $i$ depends on the entire configuration $\mathbf{S}$ of all the other spins. In serial programming, one only needs to integrate the equations of motion for individual spins in a certain order, determined by the STD decomposition (24), one equation at a time. However, Eq. (24) prohibits performing multiple spin operations at the same time, in effect prohibiting the parallelization of the integration algorithm.

Parallel integration of spin equations of motion is still possible if interaction between the spins has a finite spatial extent (Ma and Woo 2009). In this case the effective exchange field acting on a spin depends only on a finite number of neighboring spins. The simulation cell can then be subdivided into separate spatial regions, where spins belonging to different regions do not interact. Spins can then be separated into noninteracting groups, and the STD can then be applied to the groups. The integration algorithm can then be parallelized between the groups, rather than between individual spins. An intrinsic part of an MD simulation program is the linked cells algorithm. A linked cell is a local spatial region, ideally suited for the parallel implementation of the STD of evolving spin operators. The parallel implementation of spin-lattice dynamics, adopted here, relies on and benefits from the linked cells decomposition of the simulation cell (Ma and Woo 2009).

## 6      Interatomic Potentials

In the sections above, we derived equations of motion and integration algorithms for spin-lattice dynamics simulations. These simulation algorithms have now been implemented in the form of a computer program SPILADY (Ma et al. 2016). The outstanding scientific challenge in spin-lattice dynamics is the development of high-fidelity interatomic potentials, suitable for modelling the microscopic dynamics of atoms and spins in magnetic materials, composed of various chemical elements. Similarly to the interatomic potentials used in molecular dynamics (Finnis 2003), an interatomic potential can have any functional form and can be parameterized in an arbitrary way, provided that it describes the physical properties that are of interest to applications. In most cases, potentials are fitted to data derived from ab initio calculations as well as to the data derived from experimental observations. There is no universal potential yet available for any material that would be able to predict energies and forces acting between the atoms in an arbitrary atomic configuration in good agreement with ab initio calculations.

In our work, we have adopted a relatively simple functional form of the many-body non-collinear spin-lattice potential. The potential can be derived from the Hamiltonian of the form

$$\mathscr{H} = \sum_i \frac{\mathbf{P}_i^2}{2m} + U(\mathbf{R}, \mathbf{S}). \tag{25}$$

In the above Hamiltonian, the potential energy $U$ of a system with $N$ magnetic atoms is written as a function of atomic coordinates and spin vectors. One can write the energy in a Taylor series in spin variables $\mathbf{S}$. Retaining only the rotationally invariant terms, we write (Dudarev and Derlet 2007)

$$U(\mathbf{R}, \mathbf{S}) = U^{(0)}(\mathbf{R}) + \sum_i U_i^{(1)}(\mathbf{R})\mathbf{S}_i^2 + \sum_{ij} U_{ij}^{(2)}(\mathbf{R})\mathbf{S}_i \cdot \mathbf{S}_j + \sum_i U_i^{(3)}(\mathbf{R})\mathbf{S}_i^4 + \cdots .$$

(26)

This makes it possible to map Eq. (25) to a nonmagnetic many-body interatomic potential complemented with a Heisenberg-Landau Hamiltonian

$$U = U_l(\mathbf{R}) - \frac{1}{2} \sum_{ij} J_{ij}(\mathbf{R})\mathbf{S}_i \cdot \mathbf{S}_j + \sum_i A_i(\mathbf{R})\mathbf{S}_i^2 + \sum_i B_i(\mathbf{R})\mathbf{S}_i^4,$$

(27)

where $J_{ij}(\mathbf{R})$ is a coordinate-dependent exchange coupling function and $A_i(\mathbf{R})$ and $B_i(\mathbf{R})$ are the coordinate-dependent Landau coefficients. For the nonmagnetic "lattice" part $U_l(\mathbf{R})$ of the potential, we have adopted the embedded atom method (EAM) (Daw and Baskes 1984) functional form and assumed that the Landau coefficients $A_i$ and $B_i$ are functions of the effective electron density. This is by no means the only way of representing the non-collinear spin-lattice potential. However, this particular functional form has at least two advantages. Firstly, it has a clear physical meaning. The Heisenberg term describes spin-spin interactions, whereas the Landau terms describe longitudinal fluctuations of $\mathbf{S}_i$. Exchange coupling and the Landau coefficients depend on atomic coordinates, which couple the spin and lattice subsystems through forces and effective magnetic fields. Secondly, the spin and lattice equations of motion have a relatively simple form, which assists the numerical implementation of the algorithm.

A remarkable property of Eq. (27) is that in the nonmagnetic limit $\mathbf{S} = \mathbf{0}$, the spin-lattice interaction potential reduces to a conventional molecular dynamics many-body potential. This poses a question about the type of data required for fitting a spin-lattice interatomic interaction potential. The above argument shows that data on magnetic as well as on nonmagnetic atomic configurations are required, as in potential (27) one can switch on and off the magnetic spin-dependent part. Experimental observations do not always provide information about magnetic and nonmagnetic properties at the same time, making ab initio calculations an invaluable and irreplaceable source of data required for fitting spin-dependent potentials. Also, ab initio calculations allow greater freedom for preconditioning, for example, through the exploration of many non-collinear magnetic configurations in a variety of atomic environments.

The functional form of the potential given by Eq. (27) does not take into account spin-orbit coupling. Spin-orbit coupling allows the transfer of angular momentum and energy from the spin to the lattice subsystem and vice versa. Perera et al. (2016) proposed a phenomenological model for the magneto-crystalline anisotropy, which models the effect of spin-orbit coupling. They add an anisotropic term to the Hamiltonian

$$\mathscr{H}_{\mathrm{aniso}} = -C_1 \sum_i \mathbf{K}_i \cdot \mathbf{S}_i - C_2 \sum_i \mathbf{S}_i^T \cdot \mathbf{\Lambda}_i \cdot \mathbf{S}_i, \tag{28}$$

where $C_1$ and $C_2$ are adjustable parameters. $\mathbf{K}_i(\mathbf{R}) = \partial \rho_i(\mathbf{R})/\partial \mathbf{R}_i$ and $\Lambda_{i,\alpha\beta}(\mathbf{R}) = \partial^2 \rho_i(\mathbf{R})/\partial R_{i\alpha} \partial R_{i\beta}$ are the coordinate-dependent functions defining the character of the on-site magnetic anisotropy, where $\rho_i(\mathbf{R})$ describes the symmetry of local environment. This approach may help treat the effect of magneto-crystalline anisotropy, although the link between (28) and the microscopic quantum-mechanical spin-orbit coupling Hamiltonian requires further analysis.

In the study of bcc-fcc-bcc phase transitions in pure iron (Ma et al. 2017), we have fitted a new spin-lattice potential using ab initio data as input. The functional form of the spin-dependent potential was assumed to be given by Eq. (27). In the process of fitting the potential, we have generated large data sets using ab initio calculations. The data included bcc, fcc, bct, rhombohedral, amorphous, and various defect structures under the constraint that the system remained entirely nonmagnetic. Using the data, we have fitted an EAM potential, which is the first term in (27). Then, we fitted the magnetic terms, separately for bcc and fcc crystal structures. Applying the umbrella sampling and adiabatic switching thermodynamic integration to spin-lattice dynamics simulations, we have evaluated the difference between the free energies of bcc and fcc phases as a function of temperature. Each free energy difference calculation referred to a particular temperature, hence avoiding the need to perform integration from 0 K, otherwise required in other simulation approaches (Lavrentiev et al. 2010).

Figure 3 shows the calculated free energy difference between the fcc and bcc phases. When the difference is positive, bcc phase is more stable, for example, at low temperatures. Otherwise, fcc phase is more stable. The curve crosses the horizontal axis at two points, near 1130 K and then again near 1600 K. These points correspond to the bcc-fcc $\alpha - \gamma$ and fcc-bcc $\gamma - \delta$ phase transitions, respectively. The predicted transition temperatures are close to the experimentally observed transition temperatures $T_{\alpha-\gamma} = 1185$ K and $T_{\gamma-\delta} = 1667$ K. The minimum free energy difference between the fcc and bcc phases is only $-2$ meV per atom. Analysis given in Lavrentiev et al. (2010) and Ma et al. (2017) shows that $\alpha$-$\gamma$-$\delta$ phase transitions in magnetic iron stem from the interplay between magnetic excitations and lattice vibrations. The free energy contribution from non-collinear magnetic fluctuations reduces the free energy difference as temperature increases. When the temperature is higher than the Curie temperature $T_C$, the long-range magnetic order vanishes although the short-range order remains (see Ma et al. (2008) and Fig. 4).

At temperatures exceeding the Curie temperature, the contribution to entropy from magnetic excitations is superseded by the contribution from lattice vibrations, and it is the balance between entropy contributions to the free energy from spin and lattice dynamics that is ultimately responsible for the occurrence of the two, $\alpha$-$\gamma$ and $\gamma$-$\delta$, rather than one, phase transitions in iron. This also illustrates the significance of taking into account spin-lattice coupling when modelling magnetic phase transitions in any real material.
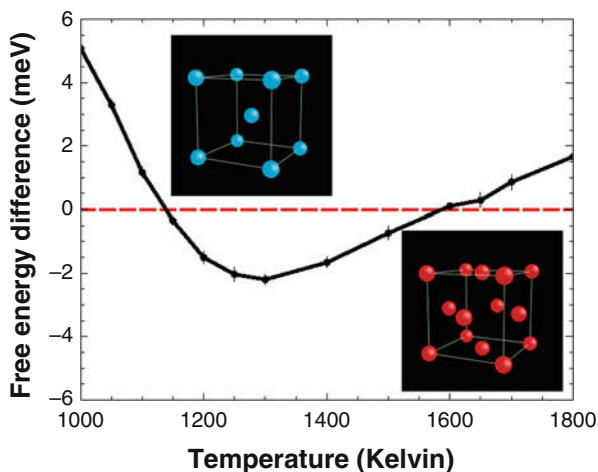
**Fig. 3** Difference between the free energies of fcc and bcc phases of magnetic iron plotted as a function of temperature. Calculations were performed using a spin-lattice potential, taken as a sum of a nonmagnetic EAM potential and a Heisenberg-Landau Hamiltonian, as detailed by Eq. (27). The minimum free energy difference between fcc and bcc phases is $-2$ meV per atom
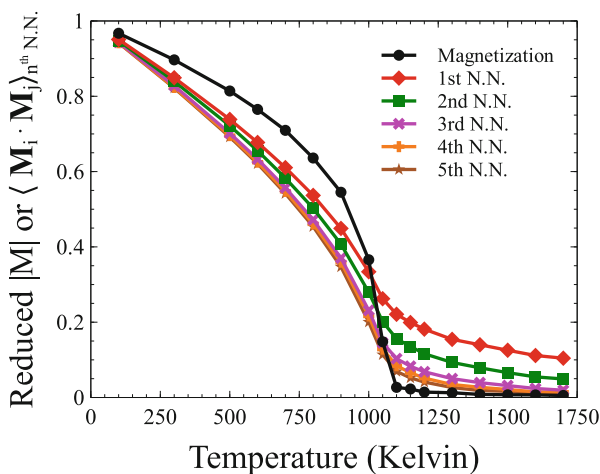


**Fig. 4** Time average projection of an atomic spin on a magnetization axis and spin-spin short-range correlation functions evaluated for the first, second, . . ., fifth nearest-neighbor atoms and plotted as functions of absolute temperature. Long-range magnetic order vanishes at the Curie temperature $T_C$, whereas the magnetic short-range order does not vanish even at temperatures well above $T_C$

# 7    Outlook and Challenges

Applications of spin-lattice dynamics are still fairly sparse. The main difficulty associated with applications of spin-lattice dynamics is that, like molecular dynamics, spin-lattice dynamics requires a sufficiently accurate many-body spin-dependent potential. Although spin-lattice dynamics has been applied extensively to simulations of various microscopic dynamic effects in pure iron, including vacancy migration (Wen et al. 2013), and magnetic excitations in cobalt (Beaujouan et al. 2012), there is still no spin-lattice potential suitable for simulating mechanical deformations, magnetic fluctuations, and defect properties at the same time. There are two main reasons why this question remains outstanding.

First, the number of degrees of freedom in spin-lattice dynamics is twice that of molecular dynamics. In molecular dynamics, one needs to fit an interatomic interaction potential to input data based on atomic configurations in $3N$ dimensions, where $N$ is the number of atoms in a simulation cell. In the case of spin-lattice dynamics, the number of degrees of freedom is $6N$ because each atom is characterized by its position as well as by the orientation and magnitude of the atomic magnetic moment. This poses a major challenge in the context of the fitting procedure as well as data generation and selection, since a significantly greater amount of data is required to span the multidimensional coordinate and spin phase space.

On the other hand, pure spin dynamics on a static lattice, which is a subset of spin-lattice dynamics, can be run as efficiently as MD. If we assume that the exchange coupling parameters $J_{ij}$ are independent of atomic coordinates, the spin-dynamics part of the integration algorithm can be used for generating detailed information about magnetic phase transitions, as illustrated in Figs. 4 and 5.
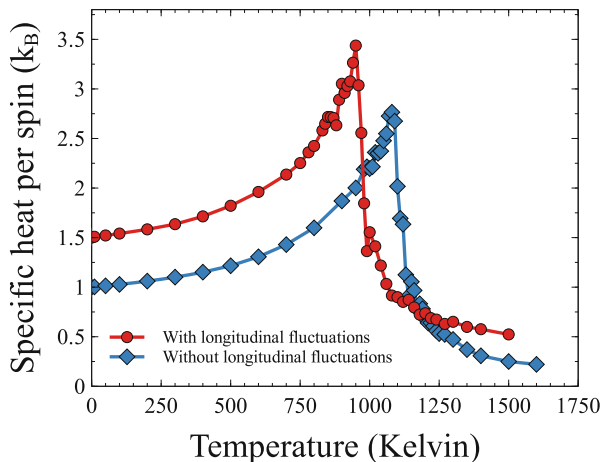


**Fig. 5** Specific heat of bcc iron evaluated using purely rotational spin dynamics with no longitudinal fluctuations included, and Langevin spin dynamics taking longitudinal fluctuations of magnetic moments into account (Ma and Dudarev 2012). The peaks correspond to second-order magnetic phase transitions, where the long-range ferromagnetic order vanishes
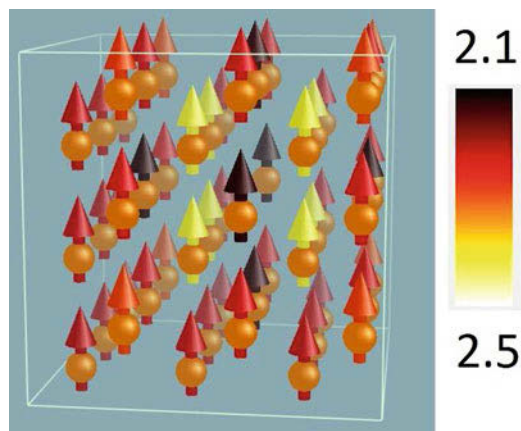
**Fig. 6** Magnetic configuration of atoms in the vicinity of a mono-vacancy in ferromagnetic bcc iron, derived from ab initio calculations. The simulation cell contains 53 atoms. The magnitude of magnetic moments is greater in the first nearest-neighbor coordination shell of the vacancy, whereas magnetic moments are suppressed in the second nearest-neighbor coordination shell. Color refers to the magnitude of magnetic moments given by the scale bar, expressed in the Bohr magneton units

Second, a suitable functional form of the spin-lattice potential is yet to be firmly established. For example, if we adopt the form given by Eq. (27), we still do not know what functional form should be used for $J_{ij}(\mathbf{R})$, $A_i(\mathbf{R})$, and $B_i(\mathbf{R})$. In Figs. 6 and 7, we show a vacancy and a $\langle 110 \rangle$ self-interstitial dumbbell configurations in bcc ferromagnetic iron. These configurations were derived from ab initio calculations. We see that the magnitude and direction of magnetic moments depend on the local environment. Atoms near a defect have significantly different magnetic moments in comparison with moments of atoms in a perfect crystal. An often used pairwise form for $J_{ij}(\mathbf{R})$ does not fit the data for defects well, although it does fit reasonably well the data on magnetic moments in a nearly perfect lattice. A good spin-lattice potential should help model magnetic configurations associated with extended defects, such as line dislocations, dislocation loops, vacancy clusters, and voids, where ab initio calculations are still impossible or too computationally demanding.

In addition to defects, alloys present an even more challenging issue. Figure 8 shows the magnetic configuration of a FeCrNi ternary alloy. The data for the figure were taken from Wróbel et al. (2015). Even though the alloy adopts a nearly perfect fcc crystal structure, its magnetic configuration is fairly complex. This implies that a spin-lattice potential must contain information about the underlying spin-dependent electronic structure to be able to reproduce magnetic properties at a reasonable level of accuracy.

The functional form of a spin-lattice potential should reflect the many-body electron interactions. A recently derived tight-binding Hamiltonian (Coury et al. 2016) for non-collinear magnetic configurations is expected to provide a good starting point for a comprehensive treatment of this problem.
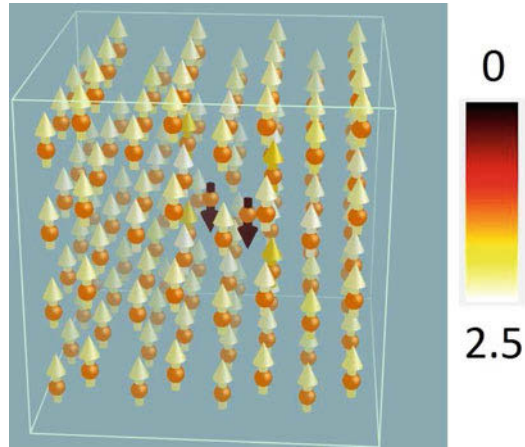
**Fig. 7** A ⟨110⟩ self-interstitial dumbbell configuration in ferromagnetic bcc iron, predicted by ab initio calculations. The supercell contains 129 iron atoms. The magnitude of magnetic moments of the two atoms forming the center of the dumbbell is significantly smaller than that of the neighboring atoms. Magnetic moments of atoms in the center of the dumbbell have antiferromagnetic orientation with respect to the ferromagnetically ordered neighboring atoms. Color represents the magnitude of an atomic magnetic moment, expressed in the Bohr magneton units

**Fig. 8** Magnetic configuration of a FeCrNi fcc alloy containing 108 atoms in the simulation cell. The alloy configuration consists of 58 Fe atoms (red), 16 Cr atoms (green), and 34 Ni atoms (blue). Color of arrows refers to the magnitude of atomic magnetic moments defined by the scale bar, expressed in the Bohr magneton units
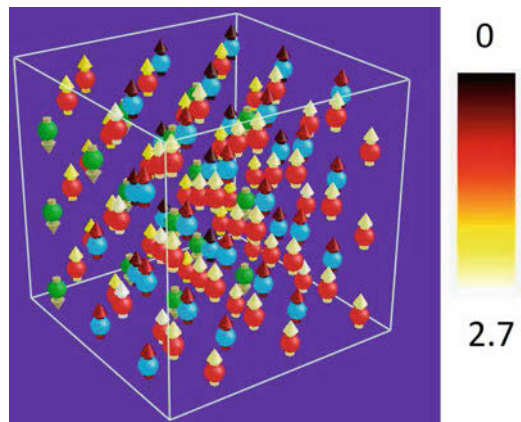


# 8 Conclusion

Spin-lattice dynamics is a powerful simulation tool for studying magnetic materials on the atomic scale. In this chapter, we have outlined the fundamental theory of spin-lattice dynamics and algorithms suitable for its numerical implementation.

Although there are still significant challenging issues that remain to be resolved, the field is open for exploration. Spin-lattice dynamics can be applied to a broad range of topics from modelling high-frequency electronic and magnetic devices to mechanical properties of magnetic alloys.

# References

Beaujouan D, Thibaudeau P, Barreteau C (2012) Anisotropic magnetic molecular dynamics of cobalt nanowires. Phys Rev B 86:174409. https://link.aps.org/doi/10.1103/PhysRevB.86.174409

Brown Jr WF (1963) Thermal fluctuations of a single-domain particle. Phys Rev 130:1677–1686. https://link.aps.org/doi/10.1103/PhysRev.130.1677

Chandrasekhar S (1943) Stochastic problems in physics and astronomy. Rev Mod Phys 15:1–89. https://link.aps.org/doi/10.1103/RevModPhys.15.1

Coury MEA, Dudarev SL, Foulkes WMC, Horsfield AP, Ma PW, Spencer JS (2016) Hubbard-like Hamiltonians for interacting electrons in $s$, $p$, and $d$ orbitals. Phys Rev B 93:075101. https://link.aps.org/doi/10.1103/PhysRevB.93.075101

Daw MS, Baskes MI (1984) Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B 29:6443–6453. https://link.aps.org/doi/10.1103/PhysRevB.29.6443

Derlet PM, Nguyen-Manh D, Dudarev SL (2007) Multiscale modeling of crowdion and vacancy defects in body-centered-cubic transition metals. Phys Rev B 76:054107. https://link.aps.org/doi/10.1103/PhysRevB.76.054107

Dudarev SL, Derlet PM (2007) Interatomic potentials for materials with interacting electrons. J Computer-Aided Mater Des 14(Suppl 1):129–140. https://doi.org/10.1007/s10820-007-9073-x, https://link.springer.com/article/10.1007/s10820-007-9073-x

Finnis MW (2003) Interatomic forces in condensed matter. Oxford series on materials modelling. Oxford University Press, Oxford

Fu CC, Willaime F, Ordejon P (2004) Stability and mobility of mono- and di-interstitials in $\alpha$-Fe. Phys Rev Lett 92:175503. https://link.aps.org/doi/10.1103/PhysRevLett.92.175503

Gilbert TL (2004) A phenomenological theory of damping in ferromagnetic materials. IEEE Trans Magn 40(6):3443–3449. https://doi.org/10.1109/TMAG.2004.836740

Hatano N, Suzuki M (2005) Finding exponential product formulas of higher orders. Springer, Berlin/Heidelberg, pp 37–68. https://doi.org/10.1007/11526216_2

Kubo R (1966) The fluctuation-dissipation theorem. Rep Prog Phys 29(1):255 http://stacks.iop.org/0034-4885/29/i=1/a=306

Landau LD, Lifshitz EM (1935) On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. Phys Z Sowjetunion 8:153–164

Lavrentiev MY, Nguyen-Manh D, Dudarev SL (2010) Magnetic cluster expansion model for bcc-fcc transitions in Fe and Fe-Cr alloys. Phys Rev B 81:184202. https://link.aps.org/doi/10.1103/PhysRevB.81.184202

Lavrentiev MY, Soulairol R, Fu CC, Nguyen-Manh D, Dudarev SL (2011) Noncollinear magnetism at interfaces in iron-chromium alloys: the ground states and finite-temperature configurations. Phys Rev B 84:144203. https://link.aps.org/doi/10.1103/PhysRevB.84.144203

Ma PW, Woo CH (2009) Parallel algorithm for spin and spin-lattice dynamics simulations. Phys Rev E 79:046703. https://link.aps.org/doi/10.1103/PhysRevE.79.046703

Ma PW, Dudarev SL (2011) Langevin spin dynamics. Phys Rev B 83:134418. https://link.aps.org/doi/10.1103/PhysRevB.83.134418

Ma PW, Dudarev SL (2012) Longitudinal magnetic fluctuations in Langevin spin dynamics. Phys Rev B 86:054416. https://link.aps.org/doi/10.1103/PhysRevB.86.054416

Ma PW, Woo CH, Dudarev SL (2008) Large-scale simulation of the spin-lattice dynamics in ferromagnetic iron. Phys Rev B 78:024434. https://link.aps.org/doi/10.1103/PhysRevB.78.024434

Ma PW, Dudarev SL, Semenov AA, Woo CH (2010) Temperature for a dynamic spin ensemble. Phys Rev E 82:031111. https://link.aps.org/doi/10.1103/PhysRevE.82.031111

Ma PW, Dudarev SL, Woo CH (2012) Spin-lattice-electron dynamics simulations of magnetic materials. Phys Rev B 85:184301. https://link.aps.org/doi/10.1103/PhysRevB.85.184301

Ma PW, Dudarev SL, Woo CH (2016) SPILADY: a parallel CPU and GPU code for spin-lattice magnetic molecular dynamics simulations. Comput Phys Commun 207(Supplement C):350–361. https://doi.org/10.1016/j.cpc.2016.05.017, http://www.sciencedirect.com/science/article/pii/S0010465516301412

Ma PW, Dudarev SL, Wróbel JS (2017) Dynamic simulation of structural phase transitions in magnetic iron. Phys Rev B 96:094418. https://link.aps.org/doi/10.1103/PhysRevB.96.094418

Nguyen-Manh D, Horsfield AP, Dudarev SL (2006) Self-interstitial atom defects in bcc transition metals: group-specific trends. Phys Rev B 73:020101. https://link.aps.org/doi/10.1103/PhysRevB.73.020101

Omelyan IP, Mryglod IM, Folk R (2001a) Algorithm for molecular dynamics simulations of spin liquids. Phys Rev Lett 86:898–901. https://link.aps.org/doi/10.1103/PhysRevLett.86.898

Omelyan IP, Mryglod IM, Folk R (2001b) Molecular dynamics simulations of spin and pure liquids with preservation of all the conservation laws. Phys Rev E 64:016105. https://link.aps.org/doi/10.1103/PhysRevE.64.016105

Omelyan IP, Mryglod IM, Folk R (2002) Construction of high-order force-gradient algorithms for integration of motion in classical and quantum systems. Phys Rev E 66:026701. https://link.aps.org/doi/10.1103/PhysRevE.66.026701

Perera D, Eisenbach M, Nicholson DM, Stocks GM, Landau DP (2016) Reinventing atomistic magnetic simulations with spin-orbit coupling. Phys Rev B 93:060402. https://link.aps.org/doi/10.1103/PhysRevB.93.060402

Pettifor DG (1995) Bonding and structure of molecules and solids. Oxford science publications, Clarendon Press. https://books.google.co.uk/books?id=r7XGPHD24fgC

Tsai SH, Krech M, Landau DP (2004) Symplectic integration methods in molecular and spin dynamics simulations. Braz J Phys 34:384–391. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-97332004000300009&nrm=iso

Tsai SH, Lee HK, Landau DP (2005) Molecular and spin dynamics simulations using modern integration methods. Am J Phys 73(7):615–624. http://doi.org/10.1119/1.1900096

Van Kampen N (2011) Stochastic processes in physics and chemistry. North-Holland Personal Library, Elsevier Science. https://books.google.co.uk/books?id=N6II-6HlPxEC

Wen H, Ma PW, Woo C (2013) Spin-lattice dynamics study of vacancy formation and migration in ferromagnetic bcc iron. J Nucl Mater 440(1):428–434. https://doi.org/10.1016/j.jnucmat.2013.05.054, http://www.sciencedirect.com/science/article/pii/S0022311513008003

Wróbel JS, Nguyen-Manh D, Lavrentiev MY, Muzyk M, Dudarev SL (2015) Phase stability of ternary fcc and bcc Fe-Cr$_{Ni}$ alloys. Phys Rev B 91:024108. https://link.aps.org/doi/10.1103/PhysRevB.91.024108

Zwanzig R (2001) Nonequilibrium statistical mechanics. Oxford University Press. https://books.google.co.uk/books?id=4cI5136OdoMC

# Part VII
# Modeling of Microstructure Evolution: Mesoscale Challenges

# Modeling of Microstructure Evolution: Mesoscale Challenges

# 48

## Marius Stan and John L. Sarrao

## Contents

### Abstract

This introductory chapter presents a perspective on multiscale modeling that emphasizes the role and challenges of mesoscale methods and their impact on understanding and predicting material properties. The predictive power of the combined experimental, theoretical, and computational mesoscale approaches is illustrated by a brief discussion of the phase field method and its application to microstructure evolution. After summarizing the main ideas of each chapter in the section, the state of the art and the future of the field are examined by asking and answering four questions: Is the 3-D representation always necessary?, Do mesoscale computational methods capture nonequilibrium?, To what degree are mesoscale methods quantitative?, and Are mesoscale methods computationally efficient?

M. Stan
Intelligent Materials Design, Applied Materials Division, Argonne National Laboratory, Argonne, IL, USA
e-mail: mstan@anl.gov

J. L. Sarrao (✉)
Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: sarrao@lanl.gov

# 1 Introduction

There is a growing appreciation in recent years of the essential role that materials phenomena at the mesoscale play in controlling bulk properties and performance (BES 2012). This is especially true for structural materials in which microstructure evolution informs deformation mechanisms as well as damage and failure. Increasingly, advanced characterization techniques are revealing microstructural variability and heterogeneity in materials, and theory and simulation have a difficult time explaining these phenomena. Recent reviews of scientific opportunities at mesoscale propose that the evolution from reductionist atomistic approaches to constructionist mesoscale strategies is representative of a broader evolution and opportunity in materials research (Sarrao 2015). Complementing other sections and their emphasis on both finer and coarser length scales, this section of the *Handbook of Materials Modeling* explores mesoscale materials phenomena and microstructure, revealing novel behavior only present at the mesoscale and describing the state of the art in modeling and computational approaches the objective of which is to explain and predict these behaviors.

This introductory chapter starts with a perspective on multiscale modeling that emphasizes the importance of mesoscale methods and the impact on understanding and predicting material properties. As an example, the phase field method, which is not the subject of a standalone chapter, is discussed in the context of "atomistic to bulk" multiscale simulation studies. The introduction continues with brief descriptions of the chapters starting with the discussion of model frontiers spanning scales in ► Chaps. 49, "Mesoscale Modeling of Dislocation-Interactions in Multilayered Materials" by Shao, S., Zhou, C., Misra, A., ► 50, "Advances in Discrete Dislocation Dynamics Simulations" by LeSar, R. and Capolungo, L. ► 51, "Mesoscale, Microstructure-Sensitive Modeling for Interface-Dominated, Nanostructured Materials" by Beyerlein, IJ. Multi-Physics and adaptive physics refinement are central to ► Chap. 52, "Adaptive Physics Refinement at the Microstructure Scale" by Germann, T.C., while ► Chap. 53, "Synchrotron Capabilities to Understand Microstructure of Additively Manufactured Materials: Challenges and Opportunities for Modeling and Simulations" by Rollett, A.D. focuses on experimental validation at the mesoscale, with specific application to APS experiments. ► Chapter 54, "Computational Modeling of Morphology Evolution in Metal-Based Battery Electrodes" by Srinivasan et al. emphasizes the technological implications of mesoscale modeling using as an example the design of Li-ion battery materials. The introductory chapter ends with a brief summary of conclusions and ideas for future studies.

# 2 Perspective on Multiscale Modeling

To understand, predict, and ultimately control matter and energy at the electronic, atomic, molecular, microstructural, and continuum levels, scientists need to investigate materials at a combination of length and time scales that are characteristic to

relevant physical and chemical phenomena. Therefore, experimental, theoretical, and computational methods must cover a wide range of space and time scales, starting with the nucleus and the electronic structure of individual or clustered atoms (Å), to nano-/microstructural features, all the way to continuum properties of the sample (cm). Along the time scale, the investigation domain ranges from excitations (ps) to nucleation of new phases (ns), all the way to diffusion (minutes, hours) and aging characteristic times (months, years).

Figure 1 shows examples of applicability domains of several theoretical and computational methods that operate at various time and length scale (Stan 2009). Some applicability limits are rooted in the physics of the associated phenomena. For example, density functional theory (DFT) – a quantum mechanical method – is best suited for investigations at short times and in small volumes, where quantum effects are prominent. Other limitations result from computational aspects such as bandwidth (speed of communication) or available computer memory. For example, molecular dynamics (MD) simulations can account for all atoms in a *mole* of matter but the time necessary to converge to solution is unrealistic, at present time. It is likely that in the next decades some methods will expand their investigation domains while others will disappear. With the accelerated advancement of theoretical and computational methodologies and capabilities, it is conceivable that quantum mechanical calculations will soon predict properties of polycrystalline, multicomponent materials, at room and higher temperatures. In recent years, considerable progress has been made in bridging mesoscale to neighboring scales by either downscaling to atomistic simulation in materials science or upscaling to the finite element simulations in structural engineering (Geers et al. 2010; Li et al. 2004). For a detailed discussion of multiscale models and simulations for soft material characterization, see Praprotnik et al. (2008).

Mesoscale interactions yield complex architectures and phenomena that serve as the building blocks of macroscopic materials behavior. Science at the mesoscale builds on dramatic advances at the atomistic and nanoscale that the research community has produced in recent years and continues to produce. The mesoscale brings profound changes, replacing the atomic granularity of matter and the quantization of energy with continuous matter and energy thereby enabling the onset of collective behavior, the interaction of coupled and competing degrees of freedom, and the appearance of defects and fluctuations that alter the behavior of perfect structures. These emergent mesoscale phenomena represent a profound challenge for multiscale materials models. A series of articles exemplify the excitement and challenges that exist at the mesoscale (Jonusauskas 2018; Sarrao and Crabtree 2012, 2015; Short and Yip 2015; Bowden et al. 1997).

# 3    The Phase Field Method

The phase field (PF) method is a powerful mesoscale simulation tool that can predict evolution of complex three-dimensional (3-D) microstructures. Applications include important materials processes such as solidification and melting,
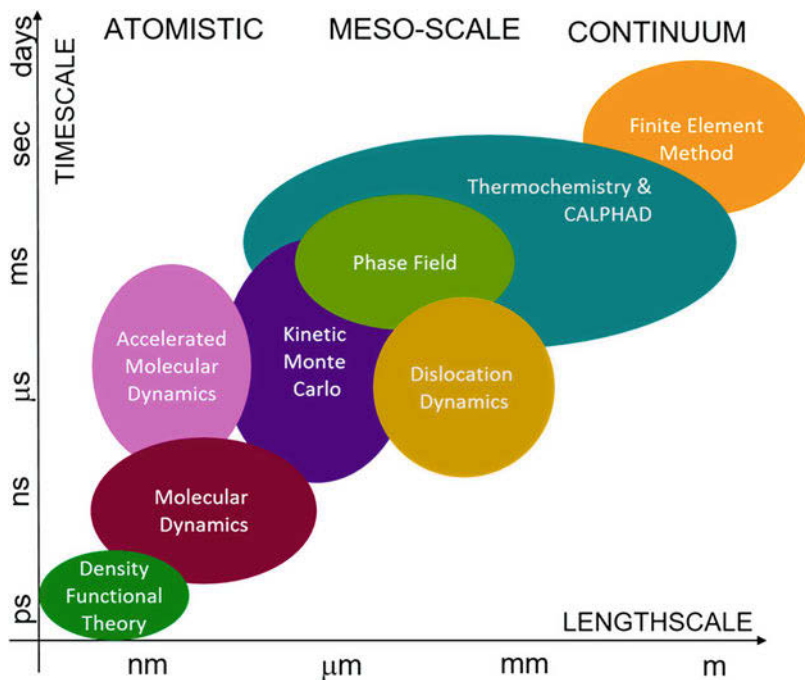
**Fig. 1** Multiscale theoretical and computational methods (Stan 2009)

ferroelectric and ferromagnetic phase transitions, phase-separation and precipitation, martensitic transformation, and twinning. The PF method requires a good understanding of the thermodynamics and kinetics of the system under examination and presents several advantages over competing methods. For example, there is no prior assumption on microstructure morphology. Also, no explicit tracking of the location of interfaces is made, which presents computational advantages. The most powerful feature is the ability to predict the evolution of multidimensional (2-D, 3-D) inhomogeneous and anisotropic microstructures with high computational efficiency. First applications of the PF method revolved around melting/solidification and solid-state phase transformations (Chen 2002; Moelans et al. 2008; Steinbach 2009; Steinbach 2013; Hu 2009; Karma and Rappel 1998). Recently, phase field simulations successfully predicted microstructure evolution of irradiated materials, including radiation-induced segregation, second-phase nucleation, void migration, interstitial loop evolution, and grain growth (Millett and Tonks 2011a, b; Li et al. 2017; Tonks et al. 2018).

In the PF method, the evolution of the microstructure is captured using two sets of field variables. One is the *concentration field*, $c$, that describes the spatial distributions of conserved material components such as chemical species. The other field variable is the *order parameter field*, $\eta$, which describes nonconserved

**Fig. 2** Schematic representation of the phase field method phase variables. The diffuse (not sharp) interface between phase 1 and adjacent phases 2 and 3 is consistent with nanoscale observations and improves computational efficiency
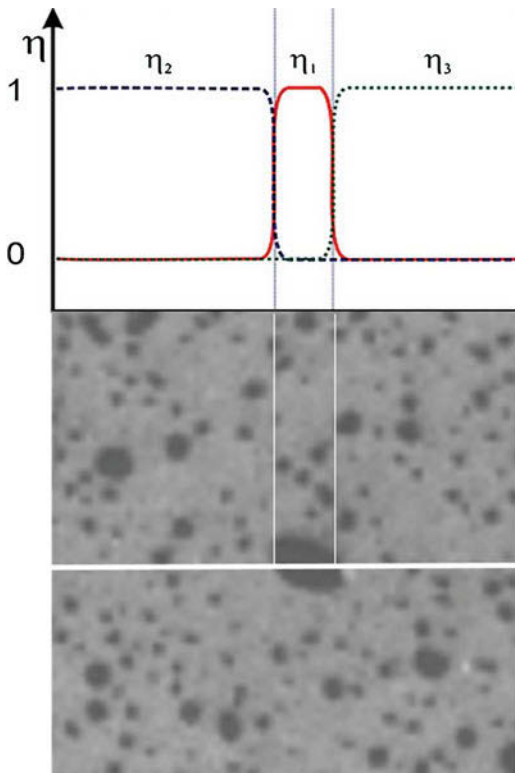
microstructural features such as crystal structures and orientations, voids or gas bubbles, or ferromagnetic domains (Fig. 2).

In the PF method, all field variables change smoothly across the interface. Therefore, the interfaces are diffuse (have a certain width). This is a key characteristic of the PF method. By avoiding singularities (locations where some inverse parameters go to infinity), the PF algorithms avoid convergence issues that are common in sharp interface and level set approaches (Ratsch et al. 2002).

We summarize here the key components of the PF mathematical formalism. For a detailed description, see Chen (2002) and Li et al. (2017).

The location and time dependence of the constituents are represented by a concentration vector $c(x,t)$, while the phases are represented by an order parameter vector $\eta(x,t)$. Here $x$ is the spatial location vector and $t$ is the time. The total free energy of the system is given by:

$$F = \int_V' \left[ f_{chem}\left(c, \eta\right) + f_{grad}\left(c, \eta\right) + f_{lr}\left(c, \eta\right) \right] dV \tag{1}$$

where $V$ is the volume of the system, $f_{chem}$ the chemical free energy density, $f_{grad}$ is the gradient energy density (interfacial energy), and $f_{lr}$ is the long-range

interaction energy density that may include elastic energy, magneto-static energy, etc. The evolution of concentration $c$ is governed by the time-dependent Cahn-Hilliard equation (Cahn 1961):

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \sum_j \left( M_{ij} \nabla \frac{\partial F}{\partial c_j} \right) + \xi_i \quad i=1,2...M \; (compositions) \tag{2}$$

The phase variable $\eta$ evolves based on the Allen-Cahn equation (Cahn and Allen 1977)

$$\frac{\partial \eta_k}{\partial t} = -L_k \frac{\partial F}{\partial \eta_k} + \xi_k \quad k=1,2...N \; (phases) \tag{3}$$

where $M_{ij}$ is the chemical mobility tensor, $L_k$ is the interface mobility, and $\xi_i$ and $\xi_k$ are thermal fluctuations. Equations (1, 2, and 3) represent a general description of the multicomponent, multiphase phase field model. Additional terms appear in the equations in PF formulations for specific applications.

Application of the PF approach to microstructure evolution has attracted growing interest because of its unique capability in predicting three-dimensional microstructure evolutions in a multicomponent and multiphase system. In the past decade, PF simulations predicted a number of important microstructure evolution phenomena and their subsequent impact on material thermo-mechanical properties of materials. In 2007, Stan *et al.* presented, for the first time, a PF model able to simulate the evolution of gas bubbles in an irradiated material (Stan et al. 2007). A more sophisticated PF model for gas bubble evolution in a polycrystalline was developed in subsequent studies (Hu et al. 2009, 2010). The approach relies on the coexistence of two phases: matrix and gas bubbles. The gas bubbles have a complex chemical composition that includes gas atoms and vacancies. In the initial approach, the grain boundaries (GBs) were fixed and did not evolve with time. Shortly after, Hu and Henager proposed a PF model that integrates elements of the kinetic Monte Carlo (kMC) method (Opplestrup et al. 2006) to simulate the evolution of void ensembles under irradiation (Hu 2009).

Besides predicting 3-D microstructure evolutions in a multicomponent and multiphase system, the PF models is able to evaluate the subsequent impact of microstructure evolution on mechanical, thermal, and magnetic properties (Klinsmann et al. 2015; Chakraborty et al. 2016).

PF simulations of polycrystalline structures with intra-granular and inter-granular pores or gas bubbles demonstrated the effect of inhomogeneous thermal conductivity on the overall thermal transport (Hu 2009; Millett et al. 2008; Millett and Tonks 2011a, b; Li et al. 2012; Chockalingam et al. 2012; Hu et al. 2015). In this case, the output microstructure resulting from PF calculations serves as input to calculations of thermal conductivity. Then, finite element simulations update the temperature field and provide feedback to new PF calculations. This iterative process can capture complex mesoscale effects and predict degradation of properties

in severe environments at very high temperatures, some close to melting (Welland et al. 2011].

## 4      Chapters in This Section

Fundamental understanding of mesoscale phenomena that influence microstructure evolution is crucial to the discovery of new materials and can accelerate material design. The chapters in this section summarize the status of a broad spectrum of mesoscale research topics and provide different perspectives regarding challenges and opportunities at mesoscale.

▶ Chapter 49, "Mesoscale Modeling of Dislocation-Interactions in Multilayered Materials" by Shao, S., Zhou, C., Misra, A. focuses on interface dislocation dynamics models for studies of ultrafine grain dynamics. The methodology relies on the atomistically informed interface dislocation dynamics (AIDD) model that is capable of bridging the mesoscale gap. The chapter includes a discussion of the fundamentals of interface deformation physics at atomic scale and the implementation into the AIDD model. To illustrate the capabilities of the AIDD model, the authors examine multilayer materials with extremely fine interface spacing exhibiting interface-dominated thermomechanical behaviors. Modeling and predicting properties of such materials on the micro and continuum scales has been challenging, with new insights recently becoming available from atomic-scale simulations.

Continuing the discussion, (▶ Chap. 50, "Advances in Discrete Dislocation Dynamics Simulations" by LeSar, R., Capolungo, L.) summarizes important, recent contributions in the field of discrete dislocation dynamics and the impact on dynamic loading/creep. The analysis goes from nanometers to tens of microns by bridging discrete dislocation dynamics simulations with harmonic transition state theory, which enables coarse graining. Advances in this methodology have taken the field closer to predicting the mechanical response of polycrystals such as textured crystalline aggregates with impurities.

▶ Chapter 51, "Mesoscale, Microstructure-Sensitive Modeling for Interface-Dominated, Nanostructured Materials" by Beyerlein, I.J. enhances the discussion of the mechanical response of polycrystalline metals and the importance of the underlying microstructure evolution and deformation mechanisms. The chapter covers recent advancements in modeling processing-microstructure-property relationships of polycrystalline metals. The model extensions highlighted in the chapter overcomes several limitations of similar approaches by incorporating nonhomogeneity. The chapter ends with recommendations for mesoscale modeling advancements as well as improvements in computational speed.

▶ Chapter 52, "Adaptive Physics Refinement at the Microstructure Scale" by Germann, T.C. focuses on long-sought goals of computational materials science and engineering: creating a simulation framework that spans all necessary length and time scales, providing the appropriate level of physics fidelity where needed, and enabling the user to trade off accuracy and computational time in an optimal manner.

After a brief review of heterogeneous multiscale methods and adaptive sampling, the authors discuss the computational workflow, data, and runtime requirements of these methods, as well as key enabling technologies such as task-based programming models, heterogeneous computer architectures, database, and machine learning algorithms.

▶ Chapter 53, "Synchrotron Capabilities to Understand Microstructure of Additively Manufactured Materials: Challenges and Opportunities for Modeling and Simulations" by Rollett, A.D. introduces the reader to techniques employed to characterize additively manufactured materials, with an emphasis on metals and ceramics. The techniques that are discussed include micro-tomography (CT), wide angle scattering (WAX), high energy diffraction microscopy (HEDM), residual stress (RS), and dynamic X-ray radiography (DXR). The chapter provides insights into both the required input for mesoscale modeling and the characterization information that can assist mesoscale model validation.

Continuing the discussion of the impact of mesoscale models on technology, ▶ Chapter 54, "Computational Modeling of Morphology Evolution in Metal-Based Battery Electrodes" by Srinivasan, V. et al. presents recent applications to lithium ion batteries. The main mesoscale challenges in developing lithium metal battery materials include dendritic growth, formation of dead lithium, and continuous decomposition of electrolyte at the solid-electrolyte interphase layer. Multiscale computational models (DFT, AIMD, kMC) capture several physical phenomena at the lithium-electrolyte interface, including mass transport, charge balance, heat transfer, and mechanical equilibrium. Successfully bridging these computational techniques is key to developing precise and accurate models.

All chapters emphasize the partnership between theory, experiment, and computation in understanding and predicting mesoscale phenomena. A new generation of supercomputers with higher computational power may also help bridge the gap between atomistic and continuum length scales by connecting with physical and chemical models of phenomena observed at the mesoscale.

## 5    Conclusions and Outlook

While acknowledging the successes of the mesoscale methods in predicting microstructure evolution and the impact on properties, a number of questions remain. For example, is the 3-D representation always necessary? There is no doubt that spatial distribution of mesoscale features is key to a number of phenomena, especially heterogeneous microstructure evolution. Furthermore, we expect that soon mesoscale computational methods will interface with 3-D material tomography to collect input data and validation information. Therefore, the answer is yes, multidimensionality is important.

Another question relates to the treatment of nonequilibrium processes that require the system to overcome energy barriers. Do mesoscale computational methods capture nonequilibrium? Of course, the laws of thermodynamics are universal and – when applied correctly – describe well both equilibrium and nonequilibrium

processes. However, the success of the mesoscale models depends upon their ability to account for nonequilibrium mechanisms of microstructure evolution, such as nucleation and growth of new phases. Capturing nucleation of secondary phases is challenging due to the small characteristic length and time scales. Often, mesoscale methods evolve the phases using as input critical nucleus information and nucleation rates from atomistic simulations such as DFT or MD. So again, coupling scales may provide the optimal path forward.

To what degree are mesoscale methods quantitative? We acknowledge that simulated microstructures are not replicas of experimentally observed microstructures but rather representations of reality. Similarly, experimental samples are not replicas but "experimental models" of real materials. Therefore, the qualitative aspect of mesoscale models is not negligible. In some instances, scientists qualitatively evaluate the dominant mechanisms behind complex phenomena as a preliminary step toward more rigorous, quantitative models and predictions. The long-term goal of mesoscale modeling is to attain a high level of precision and accuracy in quantitatively representing real materials.

The computational efficiency of mesoscale methods is also under debate. As with any computational method, going beyond the limits of applicability requires coupling with approaches that are valid at adjacent length or time scales. For example, the spatial and temporal correlations in phase field simulations require information exchange with atomistic and continuum methods. In addition, adaptive time and mesh refinement can improve the precision but decrease the numerical efficiency of solving PDF equations. Furthermore, the smallest time step and grid size are the limiting factors. All microstructure evolution simulations require highly scalable methods, as discussed in (▶ Chap. 52, "Adaptive Physics Refinement at the Microstructure Scale" by Germann, T.).

# References

BES (2012) From quanta to the continuum: opportunities for meso-scale science, BES report, Department of Energy

Bowden N, Terfort A, Carbeck J et al (1997) Self-assembly of mesoscale objects into ordered two-dimensional arrays. Science 276:233–235

Cahn JW (1961) On spinodal decomposition. Acta Metall Mater 9:795–801

Cahn JW, Allen SM (1977) A microscopic theory of domain wall motion and its experimental verification in Fe-Al alloy domain growth kinetics. J Phys Colloq 38:C7–C51

Chakraborty P, Zhang YF, Tonks MR (2016) Multi-scale modeling of microstructure dependent intergranular brittle fracture using a quantitative phase-field based method. Comput Mater Sci 113:38–52

Chen LQ (2002) Phase-field models for microstructure evolution. Annu Rev Mater Res 32:113–140

Chockalingam K, Millett PC, Tonks MR (2012) Effects of intergranular gas bubbles on thermal conductivity. J Nucl Mater 430:166–170

Geers MGD, Kouznetsova VG, Brekelmans WAM (2010) Multi-scale computational homogenization: trends and challenges. J Comput Appl Math 234:2175–2182

Hu S, Henager CH Jr. (2009) Phase-field modeling of void lattice formation under irradiation. J Nucl Mater 394:155–159

Hu S et al (2009) Phase-field modeling of gas bubbles and thermal conductivity evolution in nuclear fuels. J Nucl Mater 392:292–300

Hu S et al (2010) Application of the phase-field method in predicting gas bubble microstructure evolution in nuclear fuels. Int J Mater Res 101:515–522

Hu SY, Casella A, Lavender CA, Senor DJ, Burkes D (2015) Assessment of effective thermal conductivity in U-Mo metallic fuels with distributed gas bubbles. J Nucl Mater 462:64–76

Jonušauskas L (2018) Optical 3D printing: bridging the gaps in the mesoscale. J Opt 20:2040–8978

Karma A, Rappel WJ (1998) Quantitative phase-field modeling of dendritic growth in two and three dimensions. Phys Rev E 57:4323–4349

Klinsmann M, Rosato D, Kamlah M, McMeeking RM (2015) An assessment of the phase field formulation for crack growth. Comput Methods Appl Mech Eng 294:313–330

Li D, Li Y, Hu S, Sun X, Khaleel M (2012) Predicting thermal conductivity evolution of polycrystalline materials under irradiation using multiscale approach. Metall Mater Trans A Phys Metall Mater Sci 43A:1060–1069

Li JH, Zhang JY, Ge W et al (2004) Multi-scale methodology for complex systems. Chem Eng Sci 59:1687–1700

Li Y, Hu SY, Sun X, Stan M (2017) A review: applications of the phase field method in predicting microstructure and property evolution of irradiated nuclear materials. npj Comput Mater 3:16

Millett PC, Tonks M (2011a) Meso-scale modeling of the influence of intergranular gas bubbles on effective thermal conductivity. J Nucl Mater 412:281–286

Millett PC, Tonks M (2011b) Phase-field simulations of gas density within bubbles in metals under irradiation. Comput Mater Sci 50:2044–2050

Millett PC, Wolf D, Desai T, Rokkam S, El-Azab A (2008) Phase-field simulation of thermal conductivity in porous polycrystalline microstructures. J Appl Phys 104:033512

Moelans N, Blanpain B, Wollants P (2008) An introduction to phase-field modeling of microstructure evolution. Calphad 32:268–294

Opplestrup T, Bulatov VV, Gilmer GH, Kalos MH, Sadigh B (2006) First-passage Monte Carlo algorithm: diffusion without all the hops. Phys Rev Lett 97:230602

Praprotnik M, Delle Site L, Kremer K (2008) Multiscale simulation of soft matter: From scale bridging to adaptive resolution. Annu Rev Phys Chem 59:545–571

Ratsch C et al (2002) Level-set method for island dynamics in epitaxial growth. Phys Rev B 65:195403

Sarrao JL (2015) Opportunities and advances in mesoscale science. Curr Opinion Solid State Mater Sci 19:201–202

Sarrao JL, Crabtree GW (2012) Opportunities for mesoscale science. MRS Bull 37:1079–1088

Sarrao JL, Crabtree GW (2015) Progress in mesoscale science. MRS Bull 40:919–922

Short MP, Yip S (2015) Materials aging at mesoscale: kinetics of thermal, stress, radiation activations. Curr Opinion Solid State Mater Sci 19:245–252

Stan M (2009) Discovery and design of nuclear fuels. Mater Today 12:20–28

Stan M et al (2007) Models and simulations of nuclear fuel materials properties. J Alloys Compd 444:415–423

Steinbach I (2009) Phase-field models in materials science. Model Simul Mater Sci Eng 17:073001

Steinbach I (2013) Phase-field model for microstructure evolution at the mesoscopic scale. Annu Rev Mater Res 43:89–107

Tonks MR, Cheniour A, Aagesen L (2018) How to apply the phase field method to model radiation damage. Comput Mater Sci 147:353

Welland MJ, Lewis BJ, Thompson WT (2011) Review of high temperature thermochemical properties and application in phase-field modelling of incipient melting in defective fuel. J Nucl Mater 412:342–349

# Mesoscale Modeling of Dislocation-Interactions in Multilayered Materials

**49**

Shuai Shao, Caizhi Zhou, Amit Misra, and Jian Wang

## Contents

S. Shao
Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, USA

C. Zhou
Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO, USA

A. Misra (✉)
Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, USA
e-mail: amitmis@umich.edu

J. Wang
Mechanical and Materials Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA

**Abstract**

Multilayer materials with extremely fine interface spacing exhibit interface-dominated thermomechanical behaviors. Modeling and predicting properties of such materials on the micro- and continuum scales have been challenging, however, despite recent elucidations of the unit mechanisms from atomistic modeling and in situ experiments. This chapter presents a mesoscale, atomistically informed interface dislocation dynamics (AIDD) model that is capable of bridging the meso length-scale gap. The fundamentals of interface deformation physics at atomic-scale, implementation into the AIDD model, and the capabilities of the model are demonstrated and discussed.

# 1    Introduction

Bulk ultrafine-grained (UFG) materials are single or multiphase polycrystals with nanoscale grain size (<1000 nm). In comparison with conventional polycrystalline materials, UFG materials may exhibit increased strength/hardness, improved toughness, reduced ductility, and enhanced diffusivity (Gleiter 1989; Suryanarayana 1995; Palumbo et al. 1990; Meyers et al. 2006). Differing from coarse-grained polycrystalline materials, the most important structural characteristic of UFG materials is a large volume fraction of boundaries which significantly alters physical, mechanical, and chemical properties of materials (Hall 1951; Petch 1953; Ashby 1970; Würschum et al. 2003). For example, the UFG materials virtually exhibit no strain hardening after an initial stage of rapid strain hardening (Koch et al. 1999; Koch 2003; Van Swygenhoven and Caro 1997; Suryanarayanan Iyer et al. 1999; Carsley et al. 1998), because the density of dislocations saturates due to dynamic recovery or due to the annihilation of dislocations into the grain boundaries (Valiev et al. 2000; Jia et al. 2001; Wang et al. 2002). Besides the boundary effects, grain size also plays a crucial role in changing the deformation mechanisms, evidenced by the breakdown of the classical Hall-Petch (H-P) relationship at grain sizes below approximately 100 nm (Suryanarayanan Iyer et al. 1999; Sanders et al. 1997; Chokshi et al. 1989).

The predominant deformation mechanisms in UFG materials change from dislocation pileups to boundary-mediated unit mechanisms with decreasing size (Yamakov et al. 2002; Was and Foecke 1996; Sproul 1996; Barnett 1993; Phillips et al. 2003; Tench and White 1984; Misra et al. 2008; Anderson and Li 1995). However, the current state-of-the-art modeling methods follow the "constituent-dominated paradigm" and neglect the defect phenomena at boundaries (Cuitiño and Ortiz 1993; Dawson and Marin 1997; Roters et al. 2010; Tomé et al. 2011).

This chapter introduces a mesoscale numerical approach incorporating interface physics, the atomistically informed interface dislocation dynamics (AIDD) method, to model and predict the deformation behavior and mechanical properties of UFG materials.

## 1.1 Microstructures of UFG Materials

Here we focus on *UFG materials* with laminated microstructure which can be processed in the form of thin films via bottom-up processes such as physical vapor deposition (PVD) (Was and Foecke 1996; Mahan 2000; Ross 1994) and electrode-position (Ross 1994) or in the form of bulk materials via top-down processes such as solid-state phase transformation (Lütjering et al. 2007; Callister and Rethwisch 2007), accumulative roll bonding (ARB) (Carpenter et al. 2012; Tsuji et al. 2003), heavily drawn in situ composites (Embury 1992), or solidification (Callister and Rethwisch 2007; Shen et al. 2005; Wang et al. 2011). Figure 1 shows typical microstructures via these fabrication paths. In epitaxial films, solid-state phase transformation, or eutectic solidification, energetically favorable crystallographic orientation relationships and interface habit planes develop naturally during synthesis (Was and Foecke 1996; Mahan 2000; Wang et al. 2011; Lutjering and Williams 2007). On the other hand, bulk materials via severe plastic deformation processes often contain mechanically driven interfaces with preferred orientation relationships in order to maintain the compatibility of plastic deformation between the adjacent layers (Carpenter et al. 2012). For example, in Cu-Nb with the Kurdjumov-Sachs orientation relationship, the interface habit planes are $\{111\}_{Cu}||\{110\}_{Nb}$ for PVD (Misra et al. 2005a; Li et al. 2012; Mara et al. 2008a; Lee et al. 2012) and $\{338\}_{Cu}||\{112\}_{Nb}$ or $\{112\}_{Cu}||\{112\}_{Nb}$ for ARB (Beyerlein et al. 2014; Carpenter et al. 2012; Lee et al. 2012).

## 1.2 Mechanical Behavior of Ultrafine Scale Materials

Unlike constituent-dominated deformation mechanisms in coarse-grained materials (Fig. 2a), interfaces in nanolaminated structures play multiple roles in determining mechanical properties: sources for nucleating plastic deformation carriers (dislocations, phase transformation bands, twins, and shear bands) (Shao et al. 2015; Wang et al. 2014a, 2009a; Van Swygenhoven et al. 2006; Mara et al. 2010), barriers for impeding the propagation of these carriers (Wang et al. 2011; Misra et al. 2005a; Li et al. 2012; Mara et al. 2008b), and preferred sites for storage, reassembly, and reaction (that may lead to recovery) of interface defects (Han et al. 2014; Zhang et al. 2014; Demkowicz et al. 2008a; Höchbauer et al. 2005).

Interface spacing, $\overline{d}$ (typically layer thickness, which is often smaller than grain size), is the key microstructural parameter in determining mechanical properties of
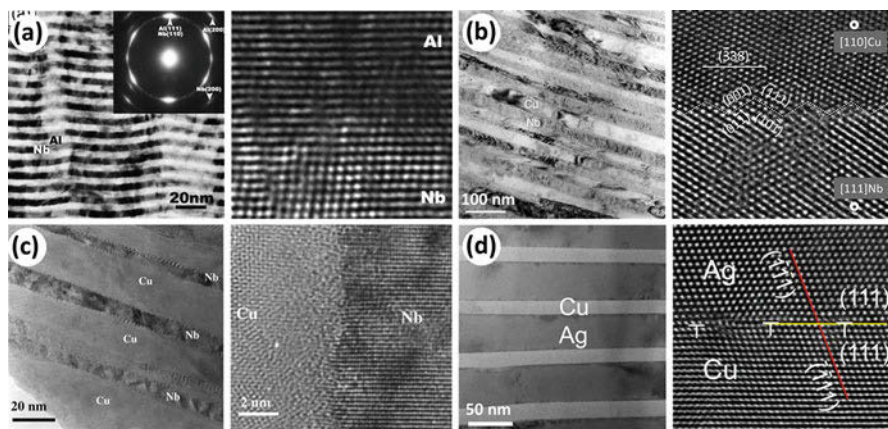
**Fig. 1** Transmission electron microscopy (TEM) and high-resolution TEM images of microstructures of various types of UFG materials. (**a**) PVD Al-Nb multilayers (Fu et al. 2008), (**b**) ARB Cu-Nb multilayers (Beyerlein et al. 2014), (**c**) wire drawn Cu-Nb nanocomposites (Sauvage et al. 2001), (**d**) rapid solidified Cu-Ag lamellar eutectic (Zheng et al. 2015)

laminated composites (Misra et al. 2005a, 2008). Assuming that a certain interface acts as barrier for continuous motion of lattice dislocations, the development of plastic deformation requires high local stresses that act on the dislocation to overcome the barrier. Corresponding to this theoretical assumption, the dislocation pileup based on Hall-Petch scaling law (Anderson and Li 1995; Misra et al. 1998; Friedman and Chrzan 1998; Huang and Spaepen 2000) is applicable at $\bar{d}$ greater than ∼100 nm, varying with properties of constituent phases. For $\bar{d}$ in the range of approximately ∼10 to ∼100 nm, dislocation pileup on the same plane is unlikely due to the strong repulsion among like-sign coplanar dislocations. The dominant deformation mechanism is confined layer slip (CLS) (Phillips et al. 2003; Anderson et al. 2003; Misra et al. 2002; Embury and Hirth 1994; Anderson et al. 1999; Freund 1990; Nix 1989) that involves propagation of single dislocation loops parallel to the interfaces within layers. At $\bar{d}$ less than approximately 10 nm, experimental data on a variety of metallic nanolaminates indicate that the hardness or strength of laminated materials (Misra et al. 1998, 2002, 2005a, 2008) shows no significant increase in the flow strength with decreasing $\bar{d}$. This behavior has been interpreted as a change in dominant deformation mechanism from CLS to interface crossing of single dislocations. In this regime, interface defect contents and the structure of line defects within interfaces strongly impact mechanical properties of UFG materials. Therefore, in addition to the nucleation of lattice dislocations, the dynamics of interface dislocations include (Fig. 2b) the motion of the interface dislocations (including climb and glide), the reactions between lattice dislocation and interface dislocations, the absorption and core spreading of lattice dislocations by interfaces of relatively low shear strength, etc.
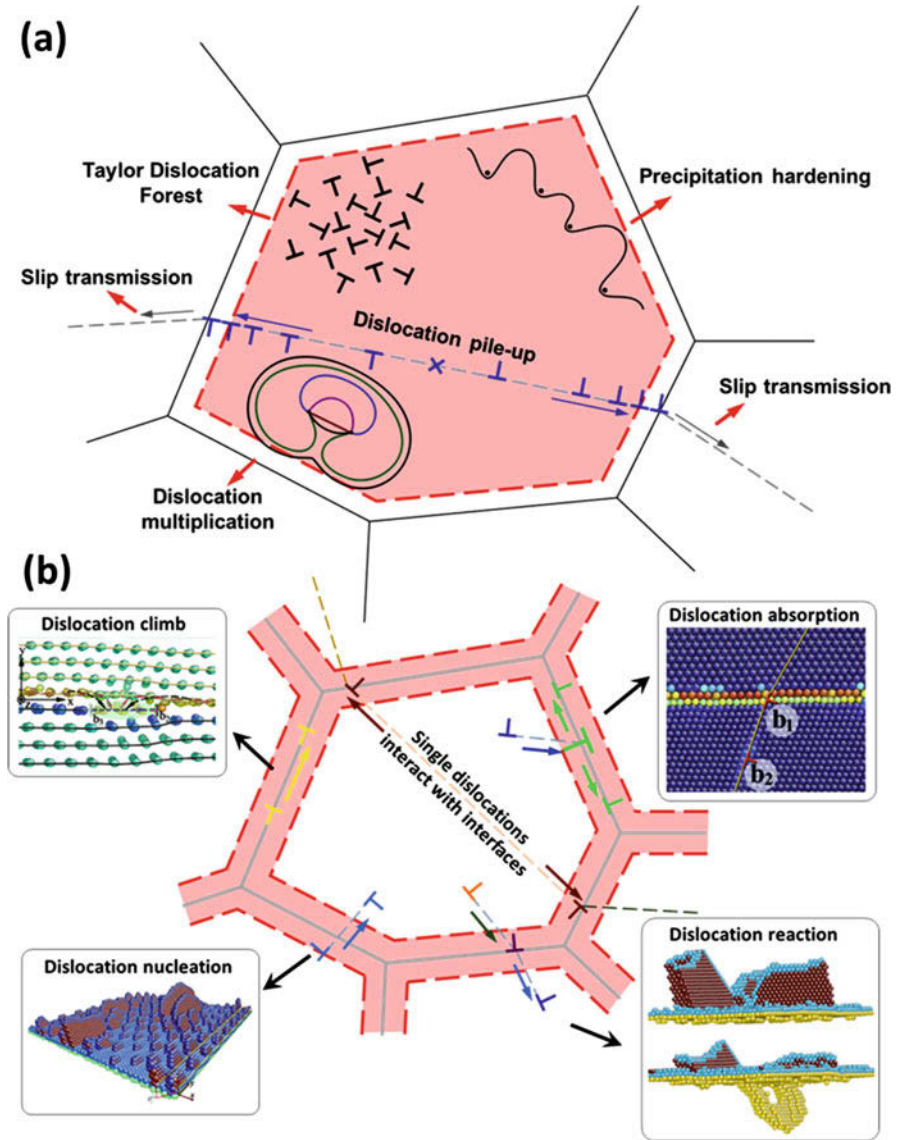
**Fig. 2** Schematic illustration of the transition in governing deformation mechanisms from constituent dominated (**a**) to interface dominated (**b**) as the characteristic length scale is decreased

## 1.3    Multiscale Perspective of Modeling Interface-Dominated Mechanical Behaviors

For interface-dominated mechanical behaviors, predictive material models must incorporate interface physics, including interface structures, unit deformation mechanisms, and dislocation density evolution at interfaces. Such interface physics is multiscale in nature. Atomistic modeling (top-left inset, Fig. 3) is able to elucidate structure and properties of interfaces and the unit processes of defect-interface interactions (Hoagland et al. 2004; Van Swygenhoven et al. 2001; Derlet and Van Swygenhoven 2002; Yamakov et al. 2003; Shenoy et al. 1999), providing the insight into the deformation processes, such as GB sliding/shear/migration (Henager et al. 2004; Wang et al. 2008a, b; Demkowicz et al. 2008b; Derlet et al. 2009), slip transmission (Sansoz and Molinari 2004; Wang and Huang 2006; Zhu et al. 2007; Afanasyev and Sansoz 2007), dislocation nucleation (Wang et al. 2009c, d, 2010, 2012a; Wang and Beyerlein 2012; Zhang et al. 2011; Kang et al. 2012a), etc. (bottom inset, Fig. 3). Beyond the temporal scale of several nanoseconds, knowledge regarding interface dislocations such as climb and glide of interface dislocations, reaction rules of interface dislocation, and evolution of interface dislocation is necessary to predict the mechanical behavior of the UFG materials. In this regard, atomistic modeling techniques are inadequate due to
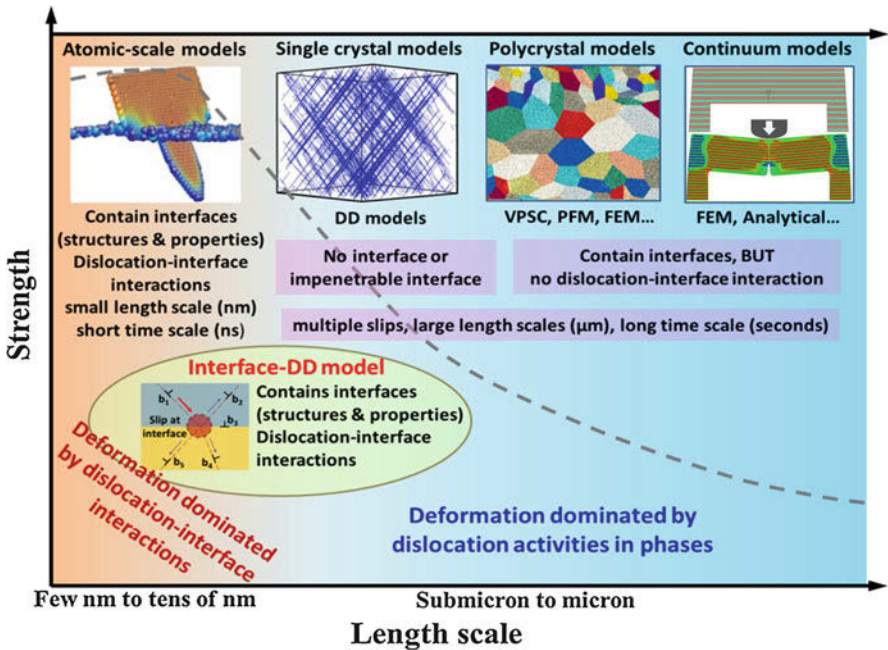


**Fig. 3** Commonly used modeling and simulation tools for studying mechanical behavior of materials at various length scales

their limitations in both length (nanometers) and temporal (nanoseconds) scales; there exists a demand for models that can close gap between atomic-scale and micro-/macroscale. Dislocation dynamics (DD) simulations (top, central-left inset, Fig. 3), in which dislocations are the simulated entities, offer a way to extend length and time scales beyond atomistic simulations and have the unique advantage of exploring dislocation activities in crystalline materials. The challenge is to describe the interactions between dislocations and interfaces. By exploiting the advantages of DD method and coupling with atomistic-level deformation mechanisms at boundaries, an atomistically informed interface dislocation dynamics (AIDD) model has been developed to link the atomistic-scale physics with macroscopic mechanical properties. The microscale grain-level crystal plasticity models (top, central-right inset of Fig. 3) can therefore benefit from the output of the AIDD model, including the constitutive relations, evolution rules of dislocations, etc. These microscale models can then inform the continuum-scale models, such as FEM and analytical constitutive models (top-right inset of Fig. 3).

This chapter reviews the fundamental aspects in developing the interface dislocation dynamics and the capabilities of the AIDD model. Conclusions and perspectives for future directions are presented.

## 2   Atomistic Information of Interface-Dominated Plasticity

Two important aspects of interface physics must be considered when describing the mechanical behavior of interface-dominated materials, namely, interfaces (structure and properties) and defect-interface interactions, both of which reside on the atomic level. Recent advances on the characterization of interface structure and properties (Kang et al. 2012a; Wang et al. 2012b, 2013, 2017; Hirth et al. 2013) as well as the elucidation of the unit interaction mechanisms between defects and interfaces (Wang et al. 2011; Shao et al. 2013; Demkowicz et al. 2012; Beyerlein et al. 2015) are achieved by combining atomistic simulations with high-resolution microscopy.

### 2.1   Atomic Structure and Thermodynamic Properties of Interfaces

In this section we categorize interfaces into four general types according to a geometrical interface classification (GIC) scheme (Wang et al. 2014a) and show that interfaces within each type share common structural as well as mechanical characteristics (Wang et al. 2014a). Within the context of the GIC scheme, we discuss these key interface-dominated plastic deformation mechanisms in detail.

The GIC scheme categorizes interfaces based on the atomic packing factors on the habit planes on the interface. For detailed discussion of the GIC scheme and information about its theoretical ground, readers are referred to ref. Wang et al. (2014a). In GIC scheme, interfaces between crystalline phases are classified into four mutually exclusive types according to the three geometric factors, namely,

compact planes (CP), compact directions (CD), and interface unit cells, which are detailed as follows:

*Type I:* the habit planes of both crystals coincide with the interface plane and the CP.

*Type II*: at least one of the two habit planes from each crystal is non-compact; however, the interface contains compact directions of the two crystals that are aligned.

*Type III*: the interface does not match the CPs or the CDs. The unit cells, however, are similar in that they have the same basic shape.

*Type IV*: the interface does not belong to Type I, Type II, or Type III.

Atomistic modeling combined with interface defect theory can be employed to determine the characteristics of all types of interfaces, such as their structures (Kang et al. 2012a, b; Wang et al. 2012b, 2014b; Misra et al. 2005b; Salehinia et al. 2015) and properties (Zheng et al. 2015; Zhu et al. 2010; Wang et al. 2009e; Liu et al. 2011; Kramer and Foecke 2002; Tian and Zhang 2012). As shown in Fig. 4, the four types of interfaces show distinct structural characteristics. In addition, interfaces that belong to the same type show strikingly similar structural features.

Two instances of *Type I* interfaces, namely, the Cu-Ni (111) interface, are shown in Fig. 4a. Indeed, the habit planes of Cu and Ni are all CPs at the interface planes;
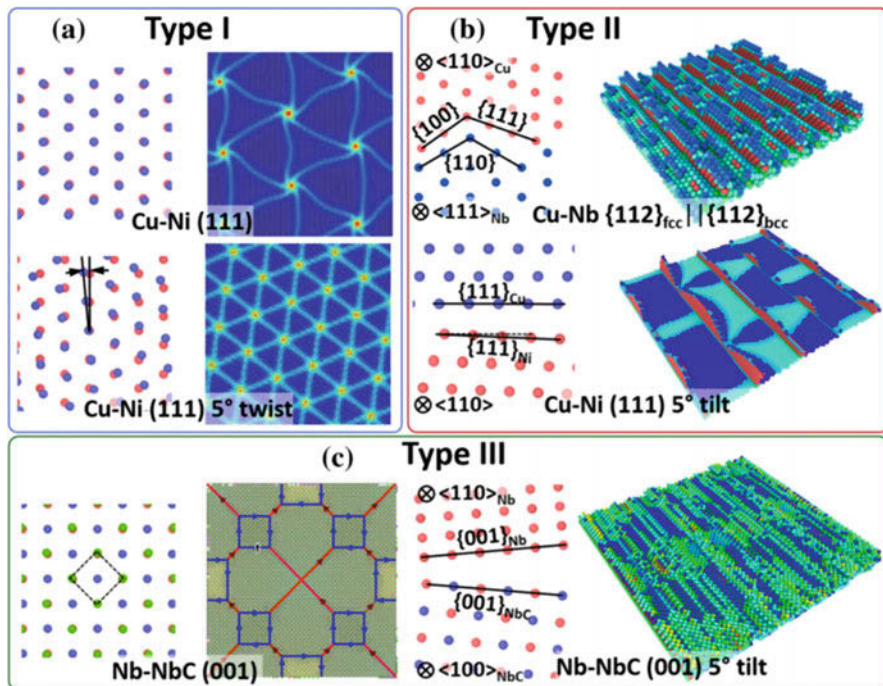


**Fig. 4** Typical examples of Type I, Type II, and Type III interfaces (Shao et al. 2018a)

the interfaces are therefore atomically flat. Networks of misfit dislocations are visible (light blue and red-colored atoms) and accommodate the transition of lattice constants, structures, and orientations (Wang et al. 2014a; Hirth et al. 2013). In general, the particular structure of the misfit dislocation network, i.e., dislocation character (edge, screw, or mixed), dislocation spacing, and number of dislocation sets, is dependent on the size and shape of the natural dichromatic patterns of the two joining crystals on the habit planes (Wang et al. 2013; Hirth et al. 2013) and the gamma surface at the interface (Salehinia et al. 2015; Chen et al. 2017; Shao et al. 2014a).

Figure 4b shows two examples of *Type II* interfaces, namely, the Cu-Nb $\{112\}_{fcc}||\{112\}_{bcc}$ and the Cu-Ni (111) interface with 5° tilt about the <110> direction. Since there is only an alignment of one CD, this type of interface is often faceted. These facets are in fact small patches of coherent regions where the respective habit planes of the joining crystals on the facet are CPs. The coherent facets are typically of relatively low excess energy. As a result, Type II interface, although comprising coherent facets separated by interfacial line defects (such as dislocations and disconnections), is still relatively low in specific energy. The Burgers vectors of the line defects can be parallel to and out-of-plane with respect to the facets as well as the interface plane.

Example of *Type III* interfaces, Nb-NbC interfaces with and without a tilt, are shown in Fig. 4c. Due to the presence of similar unit cells on the habit planes, this type of interfaces may still be preferred, even in the absence of aligning CD and CP. The non-compact habit planes are often either atomically flat (Nb-NbC interface) with relatively high atomic packing factor or faceted with patches of flat regions (Nb-NbC interface with tilt). In both scenarios, one or more sets of interface misfit dislocations/disconnections (the Burgers vectors of which can be parallel or out-of-plane with respect to the interface) separate these patches of coherent regions. Due to the strong Nb-C bonds, the *Type III* (001) interfaces are energetically preferred for Nb-NbC over other interface types.

*Type IV* interfaces typically constitute the general grain boundaries and are geometrically more complex. Characters in such interfaces can be approximately described by treating them as a superposition between small patches of interfaces of *Types I–III* and periodic interface dislocations, disconnections, and/or low-energy ledges.

One or more sets of interface line defects commonly exist on all types of interface described above (Hirth et al. 2013). Although the net defect content obeys the Frank-Bilby (F-B) theory, specific variations in the interface structures may be complex and must be resolved by atomistically informed Frank-Bilby theory and high-resolution microscopy (Wang et al. 2013; Hirth et al. 2013; Salehinia et al. 2015).

Mechanical responses (such as shear) of interfaces within each type also show similar behaviors. The shear responses of *Types I* and *II* interfaces are summarized in Fig. 5, using the *Type I* Cu-Nb Nishiyama-Wassermann (N-W, $\{111\}_{fcc}||\{110\}_{bcc}$) interface and the *Type II* Kurdjumov-Sachs (K-S, $\{112\}fcc||\{112\}bcc$) as examples. As shown, the shear strength of the *Type I* interface (Fig. 5b) is (i) significantly lower than the theoretical shear strengths in perfect crystals of Cu and Nb and (ii) is in-plane anisotropic. The shear mechanism for the Type I interface, as revealed
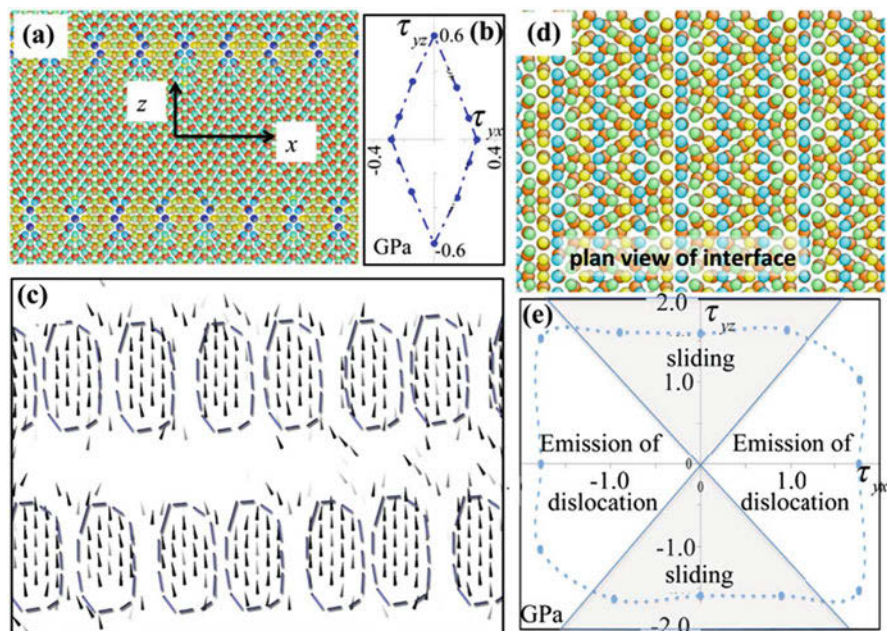
**Fig. 5** Structure, shear strengths, and shear mechanisms of Type I N-W Cu-Nb (**a–c**) interface and Type II K-S interface (**d**, **e**) (Wang et al. 2014a)

by interfacial disregistry analysis (Fig. 5c), is the easy nucleation and expansion of interface dislocations (loops) at the interface dislocation nodes or the glide of misfit dislocations if the habit planes for both joining phases are compact/near-compact. The in-plane, anisotropic shear resistance for the *Type II* interface is plotted in Fig. 5e. As shown, the shear strength along the compact directions is relatively low (Fig. 5d), and when shear deformation is applied along the non-compact directions, lattice dislocations may nucleate and emit from interfaces (Fig. 4b). Type III interfaces do not possess compact directions or compact planes; hence they typically do not slide in response to in-plane shear load. Instead, the shear loading is accommodated by nucleation and emission of lattice dislocations from interfacial intrinsic dislocations. The shear response of *Type IV* interfaces is more complicated due to the disordered atomic structure. At room temperature, they are typically associated with a high shear resistance (>2.0 GPa) in concomitant to either atomic reconstruction or the nucleation/emission of lattice dislocations.

## 2.2 Dislocation Nucleation at Interface

Lattice dislocations can be nucleated from interface via two distinct mechanisms, namely, (1) creation of a complete new loop due to stress concentration and (2) dissociation of a pre-existing misfit dislocation into a lattice dislocation half

loop and an interfacial residual dislocation. In the first mechanism, the interfacial dislocations typically retain their Burgers vector, unless the deposited portion of the nucleated loop reacts with the intrinsic interface dislocations. In the second mechanism, the Burgers vectors of the intrinsic interfacial dislocations are altered. For instance, on a pristine K-S $\{111\}_{fcc}||\{110\}_{bcc}$ interface, there are only misfit dislocations with in-plane Burgers vectors. Therefore, lattice dislocations with out-of-plane Burgers vectors can only be produced in full loops via stress concentration (Fig. 5a). On the other hand, the faceted K-S $\{112\}_{fcc}||\{112\}_{bcc}$ interfaces have lattice misfit with components both inside and out of the interface plane. Under a mechanical excitation, the intrinsic interface defects can therefore dissociate into a lattice dislocation and an interface residual.

The theoretical model and the atomistic simulations (Zhang et al. 2011, 2012; Wang et al. 2012b; Beyerlein et al. 2013a, b) indicated that only higher ranking slip systems can be nucleated from the interface. Nucleation of lattice dislocations with low-ranked or zero Schmid factors (also known as the non-Schmid effect) is not observed. The nucleation behavior of lattice dislocations from interface, similar to the interfaces' shear responses, can be classified by their types (Fig. 6). On a *Type*
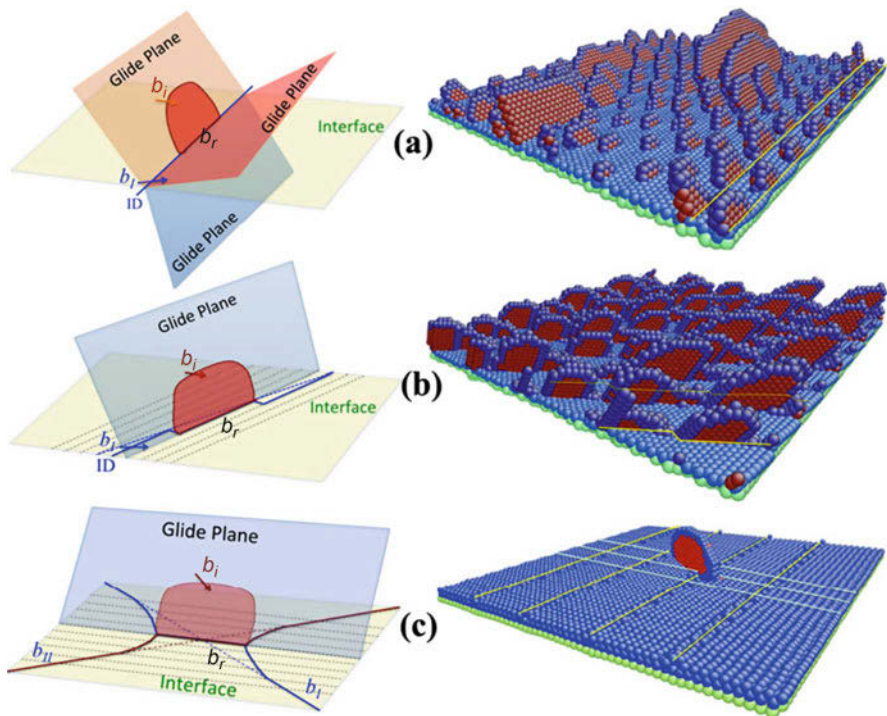


**Fig. 6** Schematic illustrations and snapshots of atomistic configurations demonstrating the mechanisms of lattice dislocation nucleation at interfaces. Three figures, respectively, show typical scenarios when interface dislocations lines are (**a**) parallel, (**b**) near-parallel, and (**c**) non-parallel to the trace of glide planes at interface (Wang et al. 2014a; Beyerlein et al. 2013b)

*I* interface (Zhang et al. 2011, 2012), there are commonly aligned traces of glide planes and the intrinsic interfacial dislocations. Due to the low shear resistance of interface, the local intensive shear results in the nucleation of lattice dislocations along the lines of interface dislocations (Fig. 6a). On *Type II* and *III* interfaces (Wang et al. 2012b, 2013), the interface dislocations may have Burgers vectors that are out of the interface plane. Their cores are typically non-planar even though they may have an in-plane Burgers vectors (Shao et al. 2018b). For these types of interfaces, dissociated-based nucleation is easier when interface dislocations are along or near-parallel (such as shown in Fig. 6a, b) to the interfacial traces of glide planes. When this parallelism cannot be established, nodes can serve as nucleation sites associated with the reassembly of interface dislocations around the node (Fig. 6c).

## 2.3    Dislocation Transmission Across Interface

Slip transmission for dislocation across interfaces takes place, to a large degree, depending on the structure of a given interface. For the investigation of slip transmission, two important criteria, both geometrical and energetic, have to be considered. Geometrically speaking, in order for transmission to happen, slip systems of the crystals on both sides of the interface have to be well aligned to provide an efficient slip path way. A good alignment between slip systems from both crystals requires that their Burgers vectors are aligned and their slip planes join at the interface. A geometrical factor, $\chi$, that describes the alignment of these slip system can therefore be expressed as (Werner and Prantl 1990)

$$\chi = \left\langle \cos\left(\frac{\pi}{2}\frac{\theta}{\theta_c}\right)\right\rangle\left\langle \cos\left(\frac{\pi}{2}\frac{\kappa}{\kappa_c}\right)\right\rangle, \tag{1}$$

where $\kappa$ is the minimum angle between their Burgers vector, $\theta$ is the minimum angle between the intersection lines that each plane makes with the interface (Roters et al. 2010), and $\theta_c$ and $\kappa_c$ are the limiting angles for $\kappa$ and $\theta$, respectively. The bracketed $<x>$ has value 0 when $x < 0$, and a value of $x$ otherwise. Apparently, an efficient transmission path across the interface would have $\chi = 1$ (Beyerlein et al. 2013b). Equation (1) also implies that when $\kappa > \kappa_c$ or $\theta > \theta_c$, $\chi = 0$, i.e., transmission is not possible. Note that $\chi$ is purely a geometrical measure and is independent of local stress. Since the Schmid factor of the outgoing slip system, $m_{\mathrm{OUT}}$, is also a limiting geometric factor (Lee et al. 1989), Eq. (1) can be modified: $\chi_\sigma = \chi m_{\mathrm{OUT}}$.

Energetically, easy transmission demands a low activation barrier. For instance, a glide lattice dislocation tends to spread within interfaces with low shear strengths (weak interfaces) and reduces its self-energy. Slip transmission would require the dislocation to re-constrict its core, leading to a relatively high energy barrier. Weak interfaces are therefore difficult for dislocations to pass. Conversely, stronger interfaces offer lower resistance for dislocation passing due to the significantly less pronounced core spreading.

Generally, since the phases joining at the interfaces can be of different crystal structure, orientation, elastic properties, chemistry, etc., dislocations are energetically favored to be trapped at interfaces rather than crossing it (Wang et al. 2008b). The complete transmission phenomenon is indeed an absorption-nucleation-emission process (Wang and Misra 2011).

## 2.4    Dynamic Recovery of Dislocations Within Interface

Remarkably, in multilayered materials fabricated via the ARB route, such as the roll-bonded Cu-Nb, the interfaces develop well-ordered atomic structures with preferred orientation relations, even under severe plastic strains (Beyerlein et al. 2014). In order to attain such extraordinary stability under such extreme conditions, mechanisms must be in place for interfaces to recover from the accumulation of damage due to the defect-interface interactions. For instance, interfacial defects can move, annihilate, or assemble then re-emit into the joining phases (Wang et al. 2011). Relative to their reaction rates, the mobility of these defects within interfaces are the limiting mechanism. Characteristics of an interface, such as its geometry, defect structure, and defect content, may all lead to different levels of difficulty for a defect to move. The interfaces with the right combinations of such attributes that facilitate motion of defects are less susceptible to accumulation of extrinsic interface defects, leading to a stable interface structure and orientation relation. These interfaces are therefore referred to as *kinetically stable interfaces*. Conversely, interfaces with poor defect mobility are referred to as *kinetically unstable interfaces*.

## 3    Fundamental Aspects of Interface Dislocation Dynamics

Interface-mediated plastic deformation mechanisms, as shown in Sect. 2, can be described by the motion/reaction of line defects at/near interfaces. By incorporating the interface physics in existing DD models, the AIDD model is developed (Wang et al. 2014a). The essential aspects are summarized as follows:

1. The interface is modeled as a piecewise flat "slip" plane, allowing the glide and climb motion of interface dislocations, depending on the relation between Burgers vector and the local interface plane normal.
2. Two distinct types of dislocations are modeled – interface dislocations and lattice dislocations. The interface dislocations include the intrinsic interface dislocations (misfit dislocations) as well as the extrinsic interface dislocations (deposited by lattice dislocations).
3. The evolution of dislocation aggregations inside the constitutive phases in the UFG and at the interfaces is dependent on the dislocation activities within phases, interfaces, and interchanges between the two. The mobility law and reaction rules of dislocations within the constitutive phases are identical to the conventional DD models (El-Awady et al. 2009; Ghoniem et al. 2000; Shao et al. 2014b; Hirth et al. 1998; Zbib and Diaz de la Rubia 2002). The laws for the motion/reaction

of the interface dislocations (e.g., drag coefficients and vacancy/interstitial migration energies) are different from that for the lattice dislocations and are, therefore, necessary to be calibrated using atomistic simulations/experimental observations.

4. Interchange between the lattice and interface dislocations can happen via the entrapment of lattice dislocation at interfaces and the nucleation/emission of lattice dislocations from interfaces.

In this section, the implementation of such interface physics, including intrinsic interface dislocation structure as well as the unit reaction mechanisms, in the AIDD model is discussed in detail.

## 3.1    Description of Interface Structures

Structure of the intrinsic interface dislocations (misfit dislocations) can be uniquely obtained via the atomistically informed Frank-Bilby (AIFB) theory (Wang et al. 2013; Hirth et al. 2013; Chen et al. 2017). Taking the Al-TiN interface as an example, the crystal structures of Al and TiN are FCC and zinc blende (in which both Ti and N atoms form FCC lattices), respectively. Experimental observations and ab initio calculations (Xie et al. 2009; Yadav et al. 2015) revealed that the adjoining Al and TiN have preferred orientation of $<110>_{Al}||<110>_{TiN}$ on $\{111\}_{Al}||\{111\}_{TiN}$ and form a *Type I* interface. The lattice constants of Al and TiN are $a_{Al} = 4.05$ Å and $a_{TiN} = 4.25$ Å, which corresponds to a 5% lattice mismatch. This mismatch results in a semi-coherent interface with the spacing between misfit dislocations of around 6 nm. Since zinc blende structure is similar to FCC, the (111) interface between Al-TiN is geometrically similar to a (111) Cu-Ni interface (Fig. 7a). Similarly, the Al-TiN interface is decorated with FCC and HCP coherent patches separated by a hexagonal/triangular network of Shockley partial dislocations (Yadav et al. 2015). The interface stacking fault energy on the Al-TiN interface, according to first principles calculations, is above 1 J/m$^2$ (Yadav et al. 2015); this energy is extremely high compared to that of the Cu-Ni (111) interface ($\sim$0.04 J/m$^2$). Atomistic simulations performed on (111) Cu-Ni interface reveals that high interface stacking fault energy results in the recombination of Shockley partial dislocations forming perfect dislocations (Fig. 7b) (Shao et al. 2013). Therefore, the misfit dislocation network of Al-TiN is modeled in AIDD as a combination of perfect and partial dislocations (Fig. 7c) (Yadav et al. 2015). In practice, to improve computational efficiency of AIDD, the spacing is enlarged while preserving the geometrical feature of the dislocation network.

## 3.2    Dislocation Nucleation at Interface

The treatment of lattice dislocation nucleation (Fig. 8a) from interfaces is carried out in several steps: (1) the dislocation lines are checked with respect to the traces of
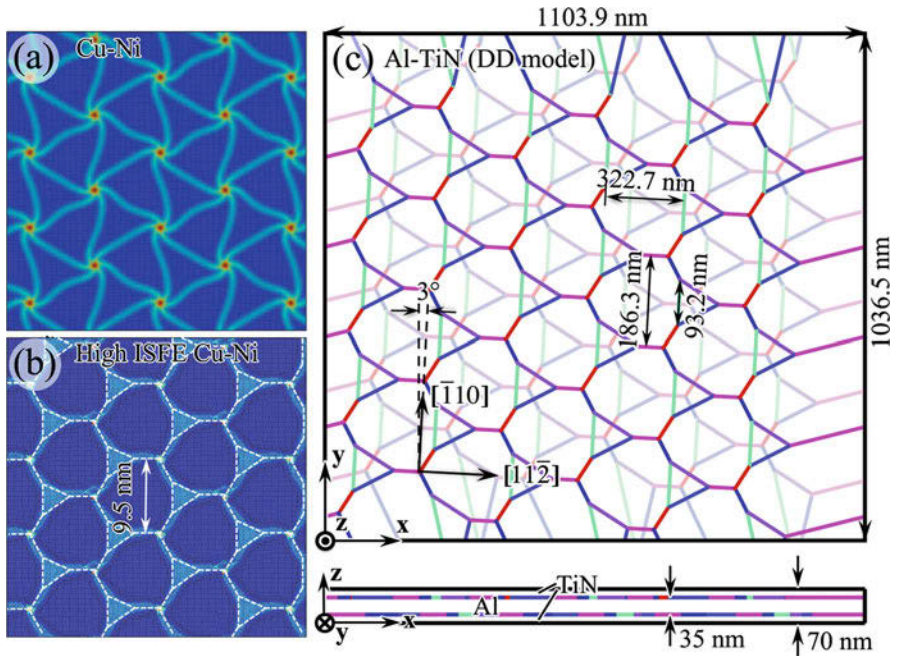
**Fig. 7** Incorporating atomistic information of the Cu-Ni interface, the dislocation structure of the Al-TiN (111) interfaces is obtained. (**a**) Shows the structure of Cu-Ni (111) interface obtained from atomistic simulation, (**b**) shows structure of an Cu-Ni (111) interface with an augmented interface stacking fault energy, (**c**) shows the dislocation structure of the Al-TiN (111) interface as implemented in the AIDD (Shao et al. 2018a)

slip planes (faint dashed lines) for parallelism, (2) a dislocation segment of a finite length that is parallel or near-parallel to a trace is "pulled" on to the trace and is registered as potential dislocation nucleation sources, and (3) the nucleation is then commenced on a fraction of the selected sources according to a global nucleation rate shown below. The various situations in which a dislocation can be selected as a source are shown in Fig. 8a.

The global dislocation nucleation rate in a crystalline material under stress can be written as (Hirth and Lothe 1982):

$$J = V C_n \omega n_0 \exp \left( \frac{-\Delta E_{line} + \pi r^2 b \tau_a / 2}{kT} \right) \tag{2}$$

where $V$ is volume of the material, $\omega$ is the frequency that an atom on the periphery joins the critical nucleus, $n_0$ is the instantaneous concentration of the number of the nucleation sites for lattice dislocation identified by step (2), and $C_n$ is a numerical term to scale the global probability. The exponential term in Eq. 2 gives the probability of finding a particular nucleation site in critical condition, $\Delta E_{line}$
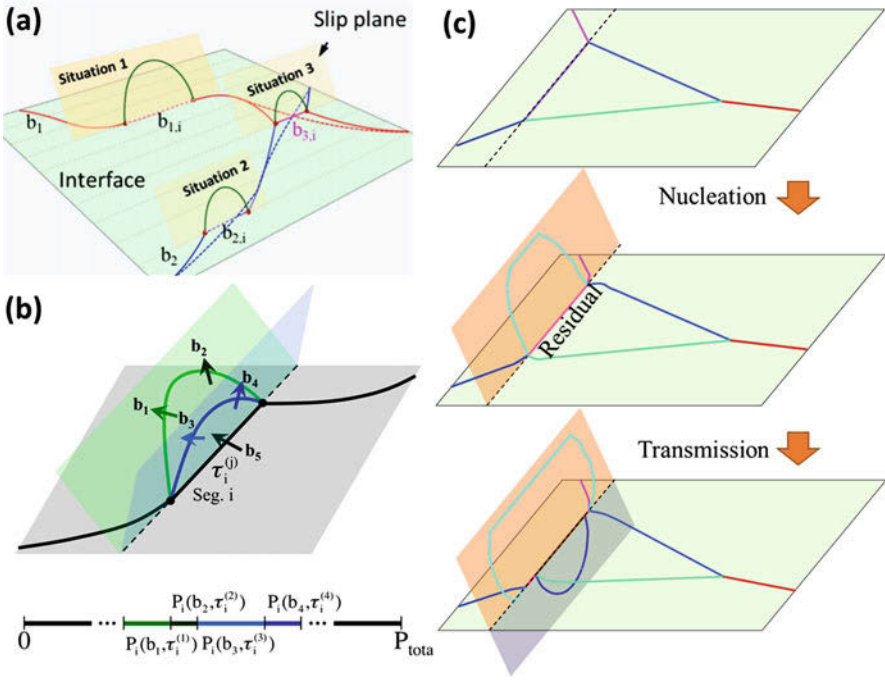
**Fig. 8** (**a**, **b**) Schematics of the implementation of nucleation mechanism in AIDD. (**c**) The realization of slip transmission across interface via a series of nucleation events, shown by actual snapshots of dislocation configurations from AIDD (Shao et al. 2018a)

is the change in dislocation line energy associated with the nucleation of a critical dislocation half loop, $r$ is the radius of a critical nucleus, and $\tau_a$ is the resolved shear stress on each of the slip systems due to the applied stress. The current version of the AIDD model does not explicitly compute contribution of stacking fault energy to nucleation. A nominal, uniform radius of critical nucleus is assumed. The actual size of the nucleus can be computed from experimental observations. This nucleation routine is performed periodically, namely, every $\Delta T = n\Delta t$ ($\Delta t$ is the numerical time step), a total of $J\Delta T$ sources are activated. The selection of the nucleation sites to be activated is achieved by a random Monte-Carlo process (Gillespie 1976) constrained by the relative nucleation probability per site (Fig. 8b).

Each of the potential nucleation sites can nucleate lattice dislocations with different Burgers vector, radius of the nucleus (half length of the interface dislocation segment) at different local resolved shear stress. Each combination of these factors is associated with a distinct probability, $P_i^{(j)}\left(b_i^{(j)}, \tau_i^{(j)}, r_i\right)$:

$$P_i^{(j)} = \exp\left(\frac{-\Delta E_i^{(j)} + \pi r_i^2 b_i^{(j)} \tau_i^{(j)}/2}{kT}\right). \tag{3}$$

Variables in Eq. 3 are similar to those in Eq. 2, but they represent local information at a particular source for a particular nucleation scenario. The Monte-Carlo selection process (Fig. 8b) is carried out by first constructing a grand probability interval according to $P_{\text{total}} = \sum_{i,j} P_i^{(j)}$ and generating a random number $\hat{r} \in [0, 1)$. A particular nucleation scenario at a nucleation source is selected if $\hat{r}P_{\text{total}}$ falls within the corresponding subinterval.

Kinematically, a lattice dislocation is nucleated by introducing a semielliptical half loop with a designated Burgers vector on to the selected slip plane. The two ends of the loop coincide with those of the interface dislocation source. Interface dislocation source is now the residual dislocation, $b_{res} = b_{\text{int.}} - b_i^{(j)}$. Under the drive of a sufficient resolved shear stress, the nucleated dislocation can propagate within the grain/phase until it reaches an adjacent interface. Slip transmission can be captured using the current nucleation scheme. Figure 8c shows the successive snapshots of dislocation structures for nucleation of a lattice dislocation and subsequent transmission of another lattice dislocation to the other side of the interface.

### 3.3   Reaction Kinematics of Dislocations Within Interface

Figure 9 lists typical reaction kinematics of interface dislocations that has been implemented into the AIDD model. An interface dislocation, depending on its Burgers vector, can glide or climb on the interface plane, i.e., when $\mathbf{b}_{inter} \cdot \hat{n}_{inter} = 0$, the dislocation is glissile; otherwise, it can only move by climb. If portions of two dislocations are in close proximity and have the same Burgers vector and opposite line sense, they annihilate (Fig. 9a). The reaction is expressed as $b_{react.\,1} - b_{react.\,2} = 0$. The reaction between two non-opposite dislocations is still permitted (Fig. 9b) if they are (1) critically close, (2) near-parallel line sense, and (3) if the resulting dislocation has lower line energy, i.e.

$$b_{react.1}^2 + b_{react.2}^2 > b_{product}^2.$$   (4)

As a result of this reaction, two new nodes form at both ends of the newly formed dislocation segment (junction). As discussed in Sect. 2.4, when the reactant dislocations are of climb type, these mechanism are responsible for the dynamic dislocation recovery at interfaces.

Near a dislocation node, more complex reactions are possible. If two dislocation "arms" branching out from a single condensed node are attracted to each other and satisfy Eq. (1), they react and result in an extended nodal structure comprising two nodes and a junction (Fig. 9c). A dislocation line (represented by the two dislocation arms colored in green in Fig. 9c) may detach from an extended node when it does not intersect other dislocation lines topologically (i.e., its arms lie adjacent to each other and on one side of the extended node). The result is an extended node with the number of arms in each condensed node reduced by one. Similar to the behavior
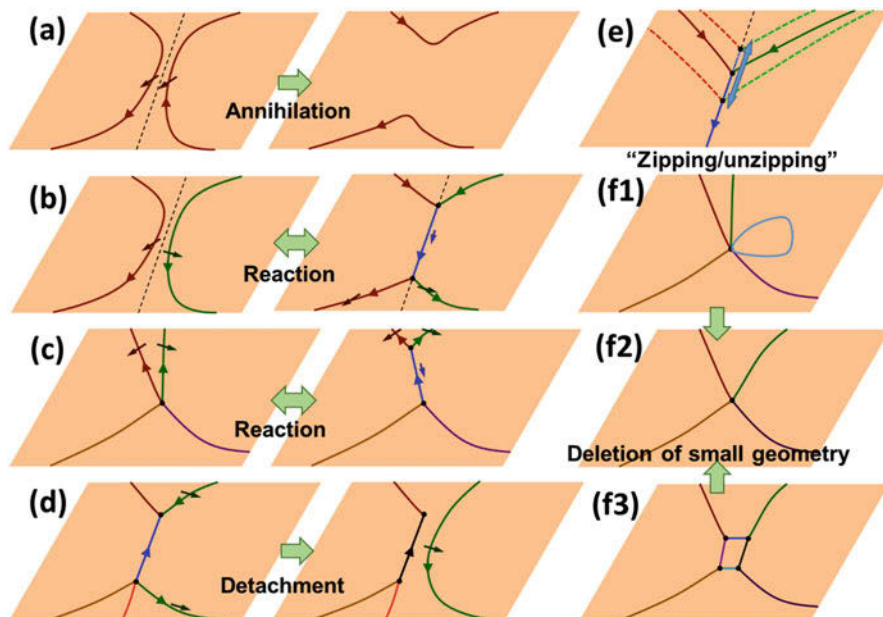
**Fig. 9** Schematic illustration of possible reaction kinematics implemented into the AIDD

of dislocation junctions in a grain/phase (El-Awady et al. 2009; Hirth et al. 1998; Zbib and Diaz de la Rubia 2002), the nodes can move via the zipping/unzipping mechanism (Fig. 9e).

The reactions among interface dislocations over an extended time and loading history may produce "debris," in the form of small dislocation geometries, such as dislocation loops and small extended nodes (Fig. 9f1, f3). Some small geometries do not have long range stress field and can be eliminated to improve computational efficiency (Fig. 9f2).

## 3.4 Dislocation-Interface Interactions

When lattice dislocations approach within a critical distance to a "strong" interface, they are captured by the interface and converted to interface dislocations; their Burgers vectors are conserved (Wang et al. 2014a; Shao et al. 2018a; Beyerlein et al. 2013a, b). If an interface dislocation with near-parallel line sense is present, the incoming lattice dislocation may react with it and form a new interface dislocation (Fig. 10a, b), once the reaction leads to a lower dislocation line energy, i.e., $b_{\text{lattice}}^2 + b_{\text{inter}}^2 > b_{\text{result}}^2$. On the other hand, when a "weak" interface with relatively low shear strength is considered, interface dislocation loop may nucleate associated with the localized shear in response to the stress field of the incoming lattice
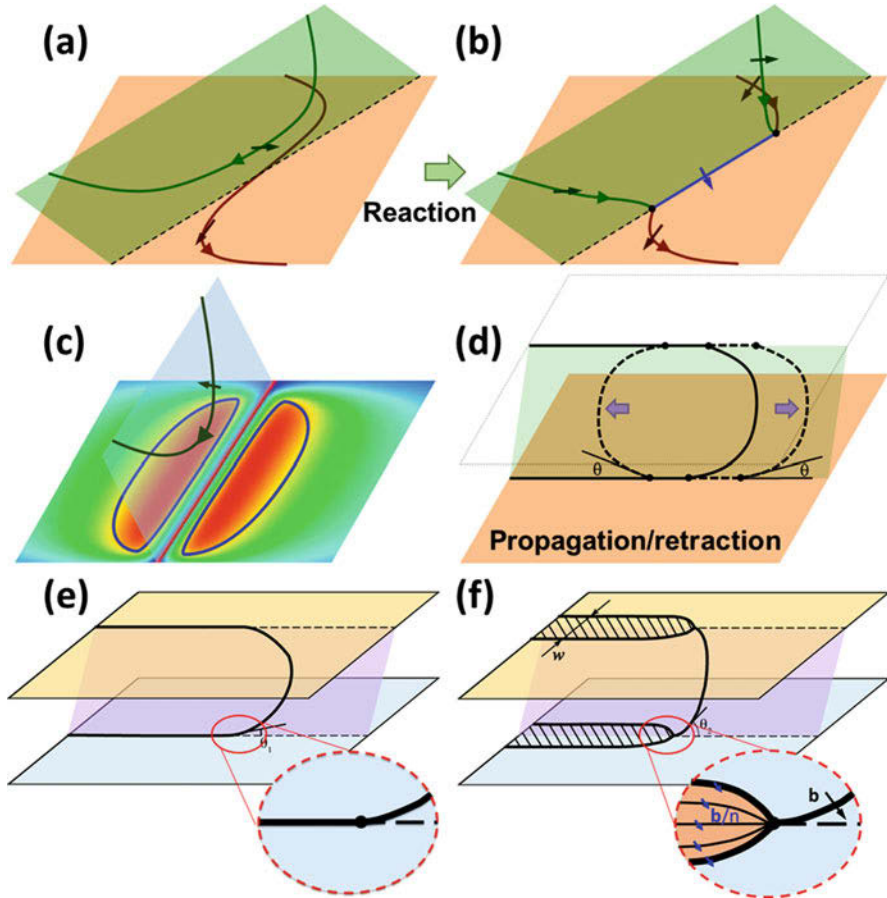
**Fig. 10** Schematic illustration of lattice dislocation – interface interaction mechanisms as implemented in the AIDD model

dislocation (Fig. 10c, the interface plane is color coded by absolute magnitude of resolved shear stress).

When the interface spacing is below ∼100 nm, dislocation motion is confined between adjacent interfaces. The CLS mechanism is modeled in AIDD (Fig. 10d) such that if the angle between a dislocation segment and the interface is small ($\theta < \theta_c$), it will be continuously deposited on the interface, and the remaining dislocation will "thread" within the layer. Conversely, if the direction of the resolved shear stress and the bow-out direction of the threading dislocation reverse, the retraction of the deposited dislocations on the interfaces (extrinsic dislocations) can be achieved.

The extrinsic dislocations create strong back stress to the threading dislocations in their vicinity and give rise to a high strain hardening rate in the multilayered

materials (Wang and Misra 2014). However, the hardening rate is dependent on the localization of interfacial dislocations' cores. Core delocalization (core-spread) is expected to moderately relieve the back stress from the extrinsic interface dislocations. A spread dislocation core is modeled in the AIDD by treating a dislocation line as a set of $n$ parallel dislocation lines distributed across a width ($w$) with reduced Burgers vector ($\mathbf{b}/n$). This is shown in Fig. 10e, f.

## 4 Applications of Atomistically Informed Interface Dislocation Dynamics

The AIDD model has been applied to the deformation of the metallic multilayers (Cu-Nb) and metal-ceramic multilayers (Al-TiN) (Wang et al. 2014a; Shao et al. 2018a; Huang et al. 2015). In this section, we briefly review these applications and discuss the capabilities of the AIDD model.

### 4.1 Metallic Multilayers

The mechanical response of Cu-Nb multilayers has been investigated by an AIDD simulation. In this simulation, the dimensions of the simulation box are $0.5 \times 0.5 \times 0.1$ µm. Equal thickness of 50 nm has been assigned to Cu and Nb layers. Full periodic boundary conditions are applied. As an initial condition, dislocation loops with random Burgers vectors (obeying crystallographic relations in Cu and Nb, respectively) were inserted in both interfaces corresponding to an initial dislocation density of $1.8 \times 10^{15}$ m$^{-2}$. An in-plane stretch loading is applied to the model by prescribing $\dot{\varepsilon}_{xx} = \dot{\varepsilon}_{yy} = \dot{\varepsilon}_{app}$ and $\sigma_{zz} = 0$, where the x and y are in-plane orthogonal directions and z is perpendicular to the interface plane. The stress-total strain curves for Cu and Nb individual layers and the multilayer are shown in Fig. 11a. Agreeing well with atomistic simulations (Abdolrahim et al. 2014), in AIDD, Cu layers yielded earlier than Nb layers, due to lower activation energy barrier in Cu. After nucleation of lattice dislocations, the lattice dislocations continued propagating similar to the Frank-Read mechanism. After continued propagation, the portion of the dislocation loop approaching the interface eventually was absorbed by the interface. Truncated by the absorption event, the remainders of the dislocation loop within the layers thread within the layers following the CLS mechanism. The gross behavior of dislocations is reflected in the evolution of dislocation densities in Fig. 11b. Corresponding to the stress-strain curves shown in Fig. 11a, an increase in dislocation density is first observed at a total strain of around 0.5% and 0.8%, corresponding to the strains at which dislocations in Cu and Nb were nucleated. The dislocation density in the two interfaces can be seen to constantly rise after nucleation. This is due to the operation of the CLS mechanism, where extrinsic interface dislocations are continuously deposited on to the interfaces as the threading dislocations propagate.
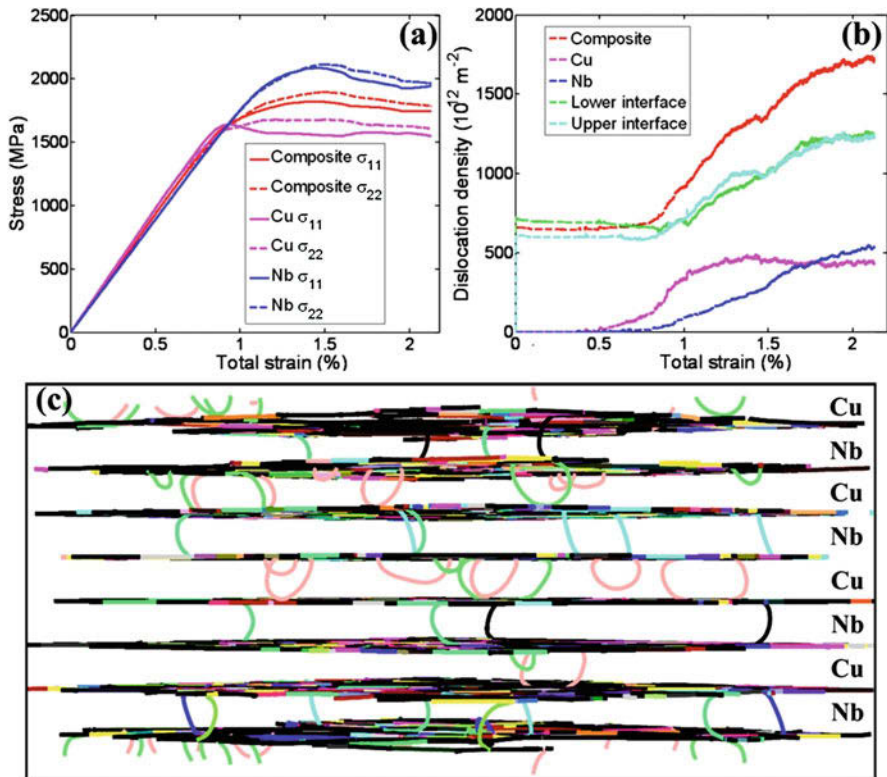
**Fig. 11** Results from the AIDD simulation of Cu-Nb multilayers, including (**a**) the stress-strain curves and (**b**) the dislocation densities for the individual layers and the composite, as well as a snapshot of dislocation configuration during simulation (Wang et al. 2014a)

## 4.2 Metal-Ceramics Multilayers

Two main sources contribute to the strain hardening of the metal-ceramic multilayers, namely, (1) the relative slip activity between the metal and ceramic layers and (2) the back stress imposed by the "deposited" (extrinsic) dislocations at interfaces (Wang and Misra 2014). The effect of the first source is demonstrated here (Shao et al. 2018a; Huang et al. 2015) using the Al-TiN as the materials of choice. The metal layers, contrasted with the ceramic layers, typically undergo much more extensive plastic deformation ascribed to factors including lower yield strength, easier nucleation of dislocations, and higher dislocation mobility. In Al layers, plastic deformation takes place following the CLS mechanism, i.e., threading dislocations propagate within the Al layer bound by the interfaces and deposit extrinsic interface dislocations. In experiments, the mechanical properties of the Al-TiN multilayers have been extensively evaluated using compressive tests normal to the interfaces. Under this loading condition, due to the aforementioned

incompatibility in plasticity and elastic constants, a large compressive (tensile) stress exists in the Al (TiN) layer parallel to the interface. The strain hardening of Al-TiN owing to the first source is ascribed to this factor, which, in turn, is dependent on the relative slip activity between the metal and ceramic layers. Next we study the deformation of two cases of Al-TiN: (1) no slip is allowed in TiN layer, and (2) moderate slip activity is allowed in TiN. The elastic constants used in the simulations are $E_{Al} = 70.7$ GPa, $\mu_{Al} = 26.2$ GPa, and $\nu_{Al} = 0.35$ and $E_{TiN} = 251$ GPa, $\mu_{TiN} = 100.4$ GPa, and $\nu_{TiN} = 0.25$. The computational cell has periodic boundary conditions in all three directions. The dimensions in x and y are around 1 μm. In the computational cell, the Al layers have the same thickness of 35 nm.

For this demonstration, both Al and TiN layers are initially dislocation free; the plasticity therefore initiates via the interface nucleation of dislocations. For the cases considered, the nucleation rate for Case 1 is zero, while nucleation rate for Case 2 is non-zero but $J_{TiN} < <J_{Al}$. The mobility of dislocations in TiN is also much lower. The results of the simulations are shown in Fig. 12. Apparently no dislocations are present in TiN in Case 1, while a few dislocations are visible in Case 2. Dislocation activity is observed in Al in both cases. In Case 2, the plastic deformation in TiN layer at total strain of 5%, $\varepsilon_p^{TiN}$, is only about 13% of the plastic strain in Al, $\varepsilon_p^{Al}$. Although this is the case, it significantly reduces the strain hardening rate (red dashed curve in Fig. 12a). The AIDD model with the plasticity enabled in TiN layer shows good agreement with the experiments by Bhattacharyya et al. (2011). The stress-strain curves normalized by the measured compressive elastic moduli of the simulations compared with the experimental results are shown in Fig. 12a. Although the yield point shows a small difference (a parameter that can be easily adjusted in AIDD), the simulated strain hardening rate for Case 2 approaches the values observed in the experiments. The results also indicate that the plastic co-deformation between the Al and TiN layers is necessary for the observed strain hardening rate in the experiment.

## 5    Remarks and Perspectives

Bulk ultrafine-grained materials, notably the multilayered materials, can exhibit a remarkable combination of thermomechanical properties, such as high yield strength, good ductility, thermal stability, and mechanical stability at extreme strains. All of which are ascribed to a number of unique characteristics of these materials, namely, high density of internal surfaces (interfaces), intergrain/interphase/interlayer geometrical constraints, highly regulated layer thickness, and finely tuned interface properties (including orientation relation, atomic and defect structure, and in-plane shear strengths).

Specifically, these extraordinary properties stem from the ability of the nanolaminates in regulating the behavior of glide dislocations. The interfaces serve multiple roles to dislocations, including the sources, barriers, as well as sites for storage, reaction, and dynamic recovery. Deep understanding toward these interface physics

**Fig. 12** Results from the AIDD simulations of Al-TiN multilayers: (**a**) stress-strain curves with experimental result from Bhattacharyya et al. (2011) and (**b**) evolution of plastic strains as a function of the total strain. Snapshots of dislocation configurations of the two simulations are shown in (**c**) for Case 1 and (**d**) for Case 2 (Shao et al. 2018a)
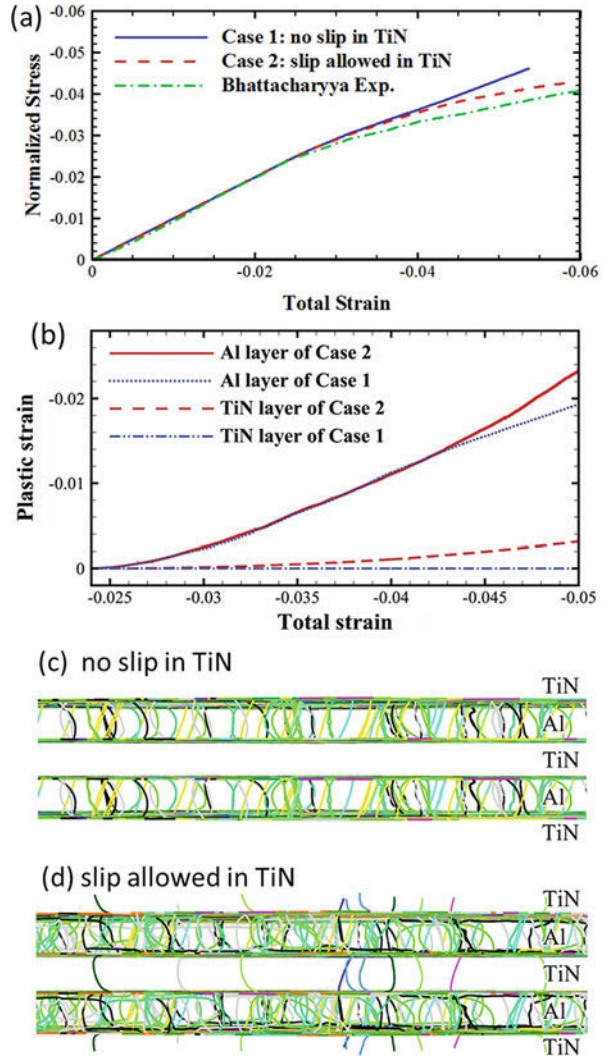
in the recent year has been provided through an abundant collection of modeling efforts at various length scales combined with experimental observations. The fruition of these theoretical studies provided insights using different techniques, including atomistic simulations (Derlet et al. 2009; Wang et al. 2008b; Wang and Misra 2011; Hoagland et al. 2002), dislocation dynamics (Zbib et al. 1998; Shehadeh et al. 2006; Overman et al. 2009), and crystal plasticity (Marin and Dawson 1998; Groh et al. 2009) simulations. However, a challenge still remains in the integration of all the discovered interface physics into a unified multiscale model to realize predictive capability (Wang et al. 2014a, 2009b).

# References

Abdolrahim N, Zbib HM, Bahr DF (2014) Multiscale modeling and simulation of deformation in nanoscale metallic multilayer systems. Int J Plast 52:33–50. https://doi.org/10.1016/j.ijplas.2013.04.002

Afanasyev KA, Sansoz F (2007) Strengthening in gold nanopillars with nanoscale twins. Nano Lett 7:2056–2062. https://doi.org/10.1021/nl070959l

Anderson PM, Li C (1995) Hall-Petch relations for multilayered materials. Nanostruct Mater 5:349–362. https://doi.org/10.1016/0965-9773(95)00250-I

Anderson PM, Foecke T, Hazzledine PM (1999) Dislocation-based deformation mechanisms in metallic nanolaminates. MRS Bull 24:27–33. https://doi.org/10.1557/S0883769400051514

Anderson PM, Bingert JF, Misra A, Hirth JP (2003) Rolling textures in nanoscale Cu/Nb multilayers. Acta Mater 51:6059–6075. https://doi.org/10.1016/S1359-6454(03)00428-2

Ashby MF (1970) The deformation of plastically non-homogeneous materials. Philos Mag 21:399–424. https://doi.org/10.1080/14786437008238426

Barnett SA (1993) Mechanic and dielectric properties advances in research and development. Academic, New York https://www.elsevier.com/books/mechanic-and-dielectric-properties/francombe/978-0-12-533017-6

Beyerlein IJ, Wang J, Zhang R (2013a) Interface-dependent nucleation in nanostructured layered composites. APL Mater 1. https://doi.org/10.1063/1.4820424

Beyerlein IJ, Wang J, Zhang R (2013b) Mapping dislocation nucleation behavior from bimetal interfaces. Acta Mater 61:7488–7499. https://doi.org/10.1016/j.actamat.2013.08.061

Beyerlein IJ, Mayeur JR, Zheng S, Mara NA, Wang J, Misra A (2014) Emergence of stable interfaces under extreme plastic deformation. Proc Natl Acad Sci 111:4386–4390. https://doi.org/10.1073/pnas.1319436111

Beyerlein IJ, Demkowicz MJ, Misra A, Uberuaga BP (2015) Defect-interface interactions. Prog Mater Sci 74:125–210. https://doi.org/10.1016/j.pmatsci.2015.02.001

Bhattacharyya D, Mara NAA, Dickerson P, Hoagland RGG, Misra A (2011) Compressive flow behavior of Al–TiN multilayers at nanometer scale layer thickness. Acta Mater 59:3804–3816. https://doi.org/10.1016/j.actamat.2011.02.036

Callister W, Rethwisch D (2007) Materials science and engineering: an introduction. https://doi.org/10.1016/0025-5416(87)90343-0

Carpenter JS, Vogel SC, Ledonne JE, Hammon DL, Beyerlein IJ, Mara NA (2012) Bulk texture evolution of Cu-Nb nanolamellar composites during accumulative roll bonding. Acta Mater 60:1576–1586. https://doi.org/10.1016/j.actamat.2011.11.045

Carsley JE, Fisher A, Milligan WW, Aifantis EC (1998) Mechanical behavior of a bulk nanostructured iron alloy. Metall Mater Trans A 29:2261–2271. https://doi.org/10.1007/s11661-998-0104-3

Chen Y, Shao S, Liu X-Y, Yadav SK, Li N, Mara N, Wang J (2017) Misfit dislocation patterns of Mg-Nb interfaces. Acta Mater 126:552–563. https://doi.org/10.1016/j.actamat.2016.12.041

Chokshi AH, Rosen A, Karch J, Gleiter H (1989) On the validity of the hall-petch relationship in nanocrystalline materials. Scr Metall 23:1679–1683. https://doi.org/10.1016/0036-9748(89)90342-6

Cuitiño AM, Ortiz M (1993) Computational modelling of single crystals. Model Simul Mater Sci Eng 1:225–263. https://doi.org/10.1088/0965-0393/1/3/001

Dawson PR, Marin EB (1997) Computational mechanics for metal deformation processes using polycrystal plasticity. Adv Appl Mech, pp 77–169. https://doi.org/10.1016/S0065-2156(08)70320-X

Demkowicz MJ, Hoagland RG, Hirth JP (2008a) Interface structure and radiation damage resistance in Cu-Nb multilayer nanocomposites. Phys Rev Lett 100:136102. https://doi.org/10.1103/PhysRevLett.100.136102

Demkowicz MJ, Wang J, Hoagland RG (2008b) Interfaces between dissimilar crystalline solids. In: Hirth JP (ed) Dislocations in solids. Elsevier B.V, Amsterdam, pp 141–205. https://doi.org/10.1016/S1572-4859(07)00003-4.

Demkowicz MJ, Misra A, Caro A (2012) The role of interface structure in controlling high helium concentrations. Curr Opin Solid State Mater Sci 16:101–108. https://doi.org/10.1016/j.cossms.2011.10.003

Derlet P, Van Swygenhoven H (2002) Length scale effects in the simulation of deformation properties of nanocrystalline metals. Scr Mater 47:719–724. https://doi.org/10.1016/S1359-6462(02)00182-3

Derlet PM, Gumbsch P, Hoagland R, Li J, McDowell DL, Van Swygenhoven H, Wang J (2009) Atomistic simulations of dislocations in confined volumes. MRS Bull 34:184–189. https://doi.org/10.1557/mrs2009.50

El-Awady JA, Wen M, Ghoniem NM (2009) The role of the weakest-link mechanism in controlling the plasticity of micropillars. J Mech Phys Solids 57:32–50. https://doi.org/10.1016/j.jmps.2008.10.004

Embury JD (1992) Micromechanical descriptions of heavily deformed materials. Scr Metall Mater 27:981–986. https://doi.org/10.1016/0956-716X(92)90460-V

Embury JD, Hirth JP (1994) On dislocation storage and the mechanical response of fine scale microstructures. Acta Metall Mater 42:2051–2056. https://doi.org/10.1016/0956-7151(94)90030-2

Freund LB (1990) The driving force for glide of a threading dislocation in a strained epitaxial layer on a substrate. J Mech Phys Solids 38:657–679. https://doi.org/10.1016/0022-5096(90)90027-2

Friedman LH, Chrzan DC (1998) Scaling theory of the Hall-Petch relation for multilayers. Phys Rev Lett 81:2715–2718. https://doi.org/10.1103/PhysRevLett.81.2715

Fu EG, Li N, Misra A, Hoagland RG, Wang H, Zhang X (2008) Mechanical properties of sputtered Cu/V and Al/Nb multilayer films. Mater Sci Eng A 493:283–287. https://doi.org/10.1016/j.msea.2007.07.101

Ghoniem N, Tong S-H, Sun L (2000) Parametric dislocation dynamics: a thermodynamics-based approach to investigations of mesoscopic plastic deformation. Phys Rev B 61:913–927. https://doi.org/10.1103/PhysRevB.61.913

Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22:403–434. https://doi.org/10.1016/0021-9991(76)90041-3

Gleiter H (1989) Nanocrystalline materials. Prog Mater Sci 33:223–315. https://doi.org/10.1016/0079-6425(89)90001-7

Groh S, Marin EB, Horstemeyer MF, Zbib HM (2009) Multiscale modeling of the plasticity in an aluminum single crystal. Int J Plast 25:1456–1473. https://doi.org/10.1016/j.ijplas.2008.11.003

Hall EO (1951) The deformation and ageing of mild steel: III discussion of results. Proc Phys Soc Sect B 64:747–753. https://doi.org/10.1088/0370-1301/64/9/303

Han WZ, Cerreta EK, Mara NA, Beyerlein IJ, Carpenter JS, Zheng SJ, Trujillo CP, Dickerson PO, Misra A (2014) Deformation and failure of shocked bulk Cu-Nb nanolaminates. Acta Mater 63:150–161. https://doi.org/10.1016/j.actamat.2013.10.019

Henager CH, Kurtz RJ, Hoagland RG (2004) Interactions of dislocations with disconnections in fcc metallic nanolayered materials. Philos Mag 84:2277–2303. https://doi.org/10.1080/14786430410001678235

Hirth JP, Lothe J (1982) Theory of dislocations, 2nd edn. Krieger Publishing Company, Malabar

Hirth JP, Zbib HM, Lothe J (1998) Forces on high velocity dislocations. Model Simul Mater Sci Eng 6:165–169

Hirth JP, Pond RC, Hoagland RG, Liu X-Y, Wang J (2013) Interface defects, reference spaces and the Frank–Bilby equation. Prog Mater Sci 58:749–823. https://doi.org/10.1016/j.pmatsci.2012.10.002

Hoagland RG, Mitchell TE, Hirth JP, Kung H (2002) On the strengthening effects of interfaces in multilayer fee metallic composites. Philos Mag A 82:643–664. https://doi.org/10.1080/01418610208243194

Hoagland RG, Kurtz RJ, Henager CH (2004) Slip resistance of interfaces and the strength of metallic multilayer composites. Scr Mater 50:775–779. https://doi.org/10.1016/j.scriptamat.2003.11.059

Höchbauer T, Misra A, Hattar K, Hoagland RG (2005) Influence of interfaces on the storage of ion-implanted He in multilayered metallic composites. J Appl Phys 98:123516. https://doi.org/10.1063/1.2149168

Huang H, Spaepen F (2000) Tensile testing of free-standing Cu, Ag and Al thin films and Ag/Cu multilayers. Acta Mater 48:3261–3269. https://doi.org/10.1016/S1359-6454(00)00128-2

Huang S, Wang J, Zhou C (2015) Effect of plastic incompatibility on the strain hardening behavior of Al-TiN nanolayered composites. Mater Sci Eng A 636:430–433. https://doi.org/10.1016/j.msea.2015.04.013

Jia D, Wang YM, Ramesh KT, Ma E, Zhu YT, Valiev RZ (2001) Deformation behavior and plastic instabilities of ultrafine-grained titanium. Appl Phys Lett 79:611–613. https://doi.org/10.1063/1.1384000

Kang K, Wang J, Zheng SJ, Beyerlein IJ (2012a) Minimum energy structures of faceted, incoherent interfaces. J Appl Phys 112:073501. https://doi.org/10.1063/1.4755789

Kang K, Wang J, Beyerlein IJ (2012b) Atomic structure variations of mechanically stable fcc-bcc interfaces. J Appl Phys 111:053531. https://doi.org/10.1063/1.3693015

Koch CC (2003) Ductility in nanostructured and ultra fine-grained materials: recent evidence for optimism. J Metastable Nanocrystalline Mater 18:9–20. https://doi.org/10.4028/www.scientific.net/JMNM.18.9

Koch CC, Morris DG, Lu K, Inoue A (1999) Ductility of nanostructured materials. MRS Bull 24:54–58. https://doi.org/10.1557/S0883769400051551

Kramer DE, Foecke T (2002) Transmission electron microscopy observations of deformation and fracture in nanolaminated Cu-Ni thin films. Philos Mag Physics Condens Matter Struct Defects Mech Prop 82:3375–3381. https://doi.org/10.1080/01418610210124478

Lee S-B, LeDonne JE, Lim SCV, Beyerlein IJ, Rollett AD (2012) The heterophase interface character distribution of physical vapor-deposited and accumulative roll-bonded Cu–Nb multilayer composites. Acta Mater 60:1747–1761. https://doi.org/10.1016/j.actamat.2011.12.007

Lee TC, Robertson IM, Birnbaum HK (1989) Prediction of slip transfer mechanisms across grain boundaries. Scr Metall 23:799–803. https://doi.org/10.1016/0036-9748(89)90534-6

Li N, Wang J, Misra A, Huang JY (2012) Direct observations of confined layer slip in Cu/Nb multilayers. Microsc Microanal 18:1155–1162. https://doi.org/10.1017/S143192761200133X

Liu Y, Bufford D, Wang H, Sun C, Zhang X (2011) Mechanical properties of highly textured Cu/Ni multilayers. Acta Mater 59:1924–1933. https://doi.org/10.1016/j.actamat.2010.11.057

Lütjering G, Williams JC (2007) Titanium: engineering materials and processes, 2nd edn. Springer, New York, pp 1–442. https://doi.org/10.1007/978-3-540-73036-1

Mahan JE (2000) Physical vapor deposition of thin films. Wiley, New York

Mara NA, Bhattacharyya D, Dickerson P, Hoagland RG, Misra A (2008a) Deformability of ultrahigh strength 5 nm Cu/Nb nanolayered composites. Appl Phys Lett 92:231901. https://doi.org/10.1063/1.2938921

Mara NA, Bhattacharyya D, Hoagland RG, Misra A (2008b) Tensile behavior of 40 nm Cu/Nb nanoscale multilayers. Scr Mater 58:874–877. https://doi.org/10.1016/j.scriptamat.2008.01.005

Mara NA, Bhattacharyya D, Hirth JP, Dickerson P, Misra A (2010) Mechanism for shear banding in nanolayered composites. Appl Phys Lett 97:021909. https://doi.org/10.1063/1.3458000

Marin EB, Dawson PR (1998) On modelling the elasto-viscoplastic response of metals using polycrystal plasticity. Comput Methods Appl Mech Eng 165:1–21. https://doi.org/10.1016/S0045-7825(98)00034-6

Meyers MAM, Mishra A, Benson DJ (2006) Mechanical properties of nanocrystalline materials. Prog Mater Sci 51:427–556. https://doi.org/10.1016/j.pmatsci.2005.08.003

Misra A, Verdier M, Lu Y, Kung H, Mitchell T, Nastasi M, Embury J (1998) Structure and mechanical properties of Cu-X (X = Nb,Cr,Ni) nanolayered composites. Scr Mater 39:555–560. https://doi.org/10.1016/S1359-6462(98)00196-1

Misra A, Hirth JP, Kung H, Kunc H, Kung H (2002) Single-dislocation-based strengthening mechanisms in nanoscale metallic multilayers. Philos Mag A 82:2935–2951. https://doi.org/10.1080/01418610208239626

Misra A, Hirth JP, Hoagland RG (2005a) Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites. Acta Mater 53:4817–4824. https://doi.org/10.1016/j.actamat.2005.06.025

Misra A, Zhang X, Hammon D, Hoagland RG (2005b) Work hardening in rolled nanolayered metallic composites. Acta Mater 53:221–226. https://doi.org/10.1016/j.actamat.2004.09.018

Misra A, Demkowicz MJ, Wang J, Hoagland RG (2008) The multiscale modeling of plastic deformation in metallic nanolayered composites. JOM 60:39–42. https://doi.org/10.1007/s11837-008-0047-6

Nix WD (1989) Mechanical properties of thin films. Metall Trans A 20:2217–2245. https://doi.org/10.1007/BF02666659

Overman NR, Overman CT, Zbib HM, Bahr DF (2009) Yield and deformation in biaxially stressed multilayer metallic thin films. J Eng Mater Technol 131:041203. https://doi.org/10.1115/1.3183775

Palumbo G, Thorpe SJ, Aust KT (1990) On the contribution of triple junctions to the structure and properties of nanocrystalline materials. Scr Metall Mater 24:1347–1350. https://doi.org/10.1016/0956-716X(90)90354-J

Petch NJ (1953) The cleavage strength of polycrystals. J Iron Steel Inst 174:25–28. https://doi.org/10.1007/BF01972547

Phillips M, Clemens B, Nix W (2003) Microstructure and nanoindentation hardness of Al/Al 3 Sc multilayers. Acta Mater 51:3171–3184. https://doi.org/10.1016/S1359-6454(03)00128-9

Ross CA (1994) Electrodeposited multilayer thin films. Annu Rev Mater Sci 24:159–188. https://doi.org/10.1146/annurev.ms.24.080194.001111

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58:1152–1211. https://doi.org/10.1016/j.actamat.2009.10.058

Salehinia I, Shao S, Wang J, Zbib HM (2015) Interface structure and the inception of plasticity in Nb/NbC nanolayered composites. Acta Mater 86:331–340. https://doi.org/10.1016/j.actamat.2014.12.026

Sanders PG, Eastman JA, Weertman JR (1997) Elastic and tensile behavior of nanocrystalline copper and palladium. Acta Mater 45:4019–4025. https://doi.org/10.1016/S1359-6454(97)00092-X

Sansoz F, Molinari JF (2004) Incidence of atom shuffling on the shear and decohesion behavior of a symmetric tilt grain boundary in copper. Scr Mater 50:1283–1288. https://doi.org/10.1016/j.scriptamat.2004.02.031

Sauvage X, Renaud L, Deconihout B, Blavette D, Ping D, Hono K (2001) Solid state amorphization in cold drawn Cu/Nb wires. Acta Mater 49:389–394. https://doi.org/10.1016/S1359-6454(00)00338-4.

Shao S, Wang J, Misra A, Hoagland RG (2013) Spiral patterns of dislocations at nodes in (111) semi-coherent FCC interfaces. Sci Rep 3. https://doi.org/10.1038/srep02448

Shao S, Wang J, Misra A (2014a) Energy minimization mechanisms of semi-coherent interfaces. J Appl Phys 116:023508. https://doi.org/10.1063/1.4889927

Shao S, Abdolrahim N, Bahr DF, Lin G, Zbib HM (2014b) Stochastic effects in plasticity in small volumes. Int J Plast 52:117–132. https://doi.org/10.1016/j.ijplas.2013.09.005

Shao S, Wang J, Beyerlein IJ, Misra A (2015) Glide dislocation nucleation from dislocation nodes at semi-coherent {111} Cu–Ni interfaces. Acta Mater 98:206–220. https://doi.org/10.1016/j.actamat.2015.07.044

Shao S, Misra A, Huang H, Wang J (2018a) Micro-scale modeling of interface-dominated mechanical behavior. J Mater Sci:5546 https://doi.org/10.1007/s10853-017-1662-9

Shao S, Akasheh F, Wang J, Liu Y (2018b) Alternative misfit dislocations pattern in semi-coherent FCC {100} interfaces. Acta Mater 144:177–186. https://doi.org/10.1016/j.actamat.2017.10.052

Shehadeh MA, Bringa EM, Zbib HM, McNaney JM, Remington BA (2006) Simulation of shock-induced plasticity including homogeneous and heterogeneous dislocation nucleations. Appl Phys Lett 89:171918. https://doi.org/10.1063/1.2364853

Shen TD, Schwarz RB, Zhang X (2005) Bulk nanostructured alloys prepared by flux melting and melt solidification. Appl Phys Lett 87:1–3. https://doi.org/10.1063/1.2056610.

Shenoy VB, Miller R, Tadmor EB, Rodney D, Phillips R, Ortiz M (1999) An adaptive finite element approach to atomic-scale mechanics—the quasicontinuum method. J Mech Phys Solids 47:611–642. https://doi.org/10.1016/S0022-5096(98)00051-9.

Sproul WD (1996) New routes in the preparation of mechanically hard films. Science 273:889–892. https://doi.org/10.1126/science.273.5277.889

Suryanarayana C (1995) Nanocrystalline materials. Int Mater Rev 40:41–64. https://doi.org/10.1179/imr.1995.40.2.41

Suryanarayanan Iyer R, Frey CA, Sastry SM, Waller B, Buhro W (1999) Plastic deformation of nanocrystalline Cu and Cu–0.2 wt.% B. Mater Sci Eng A 264:210–214. https://doi.org/10.1016/S0921-5093(98)01027-2

Tench D, White J (1984) Enhanced tensile strength for electrodeposited nickel-copper multilayer composites. Metall Trans A 15:2039–2040. https://doi.org/10.1007/BF02646838

Tian YZ, Zhang ZF (2012) Bulk eutectic Cu-Ag alloys with abundant twin boundaries. Scr Mater 66:65–68. https://doi.org/10.1016/j.scriptamat.2011.09.024

Tomé CN, Beyerlein IJ, Wang J, McCabe RJ (2011) A multi-scale statistical study of twinning in magnesium. JOM 63:19–23. https://doi.org/10.1007/s11837-011-0038-x

Tsuji N, Saito Y, Lee SH, Minamino Y (2003) ARB (accumulative roll-bonding) and other new techniques to produce bulk ultrafine grained materials. Adv Eng Mater 5:338–344. https://doi.org/10.1002/adem.200310077

Valiev RZ, Islamgaliev RK, Alexandrov IV (2000) Bulk nanostructured materials from severe plastic deformation. Prog Mater Sci 45:103–189. https://doi.org/10.1016/S0079-6425(99)00007-9

Van Swygenhoven H, Caro A (1997) Molecular dynamics computer simulation of nanophase Ni: structure and mechanical properties. Nanostruct Mater 9:669–672. https://doi.org/10.1016/S0965-9773(97)00147-5

Van Swygenhoven H, Caro A, Farkas D (2001) A molecular dynamics study of polycrystalline fcc metals at the nanoscale: grain boundary structure and its influence on plastic deformation. Scr Mater 309–310:1513–1516. https://doi.org/10.1016/S0921-5093(00)01794-9

Van Swygenhoven H, Derlet PM, Frøseth a G (2006) Nucleation and propagation of dislocations in nanocrystalline fcc metals. Acta Mater 54:1975–1983. https://doi.org/10.1016/j.actamat.2005.12.026

Wang J, Beyerlein IJ (2012) Atomic structures of symmetric tilt grain boundaries in hexagonal close packed (hcp) crystals. Model Simul Mater Sci Eng 20:024002. https://doi.org/10.1088/0965-0393/20/2/024002

Wang J, Huang H (2006) Novel deformation mechanism of twinned nanowires. Appl Phys Lett 88:203112. https://doi.org/10.1063/1.2204760

Wang J, Misra A (2011) An overview of interface-dominated deformation mechanisms in metallic multilayers. Curr Opin Solid State Mater Sci 15:20–28. https://doi.org/10.1016/j.cossms.2010.09.002

Wang J, Misra A (2014) Strain hardening in nanolayered thin films. Curr Opin Solid State Mater Sci 18:19–28. https://doi.org/10.1016/j.cossms.2013.10.003

Wang YM, Ma E, Chen MW (2002) Enhanced tensile ductility and toughness in nanostructured Cu. Appl Phys Lett 80:2395–2397. https://doi.org/10.1063/1.1465528

Wang J, Hoagland RG, Hirth JP, Misra A (2008a) Atomistic simulations of the shear strength and sliding mechanisms of copper–niobium interfaces. Acta Mater 56:3109–3119. https://doi.org/10.1016/j.actamat.2008.03.003

Wang J, Hoagland RG, Hirth JP, Misra A (2008b) Atomistic modeling of the interaction of glide dislocations with "weak" interfaces. Acta Mater 56:5685–5693. https://doi.org/10.1016/j.actamat.2008.07.041

Wang J, Hirth JP, Tomé CN (2009a) (over(1, ¯) 0 1 2) Twinning nucleation mechanisms in hexagonal-close-packed crystals. Acta Mater 57:5521–5530. https://doi.org/10.1016/j.actamat.2009.07.047.

Wang J, Hoagland RGG, Misra A (2009b) Mechanics of nanoscale metallic multilayers: from atomic-scale to micro-scale. Scr Mater 60:1067–1072. https://doi.org/10.1016/j.scriptamat.2008.11.035

Wang J, Hirth JP, Tomé CN (2009c) (1¯012) Twinning nucleation mechanisms in hexagonal-close-packed crystals. Acta Mater 57:5521–5530. https://doi.org/10.1016/j.actamat.2009.07.047

Wang J, Hoagland RG, Hirth JP, Capolungo L, Beyerlein IJ, Tomé CN (2009d) Nucleation of a (1012) twin in hexagonal close-packed crystals. Scr Mater 61:903–906. https://doi.org/10.1016/j.scriptamat.2009.07.028

Wang J, Hoagland RG, Misra A (2009e) Room-temperature dislocation climb in metallic interfaces. Appl Phys Lett 94:131910. https://doi.org/10.1063/1.3111137

Wang J, Li N, Anderoglu O, Zhang X, Misra A, Huang JY, Hirth JP (2010) Detwinning mechanisms for growth twins in face-centered cubic metals. Acta Mater 58:2262–2270. https://doi.org/10.1016/j.actamat.2009.12.013

Wang J, Beyerlein IJ, Mara NA, Bhattacharyya D (2011) Interface-facilitated deformation twinning in copper within submicron Ag-Cu multilayered composites. Scr Mater 64:1083–1086. https://doi.org/10.1016/j.scriptamat.2011.02.025

Wang J, Misra A, Hoagland RG, Hirth JP (2012a) Slip transmission across fcc/bcc interfaces with varying interface shear strengths. Acta Mater 60:1503–1513. https://doi.org/10.1016/j.actamat.2011.11.047

Wang J, Kang K, Zhang RF, Zheng SJ, Beyerlein IJ, Mara NA (2012b) Structure and property of interfaces in ARB Cu/Nb laminated composites. JOM 64:1208–1217. https://doi.org/10.1007/s11837-012-0429-7

Wang J, Zhang R, Zhou C, Beyerlein IJ, Misra A (2013) Characterizing interface dislocations by atomically informed Frank-Bilby theory. J Mater Res 28:1646–1657. https://doi.org/10.1557/jmr.2013.34

Wang J, Zhou C, Beyerlein IJIJ, Shao S (2014a) Modeling interface-dominated mechanical behavior of nanolayered crystalline composites. JOM 66:102–113. https://doi.org/10.1007/s11837-013-0808-8

Wang J, Zhang RF, Zhou CZ, Beyerlein IJ, Misra A (2014b) Interface dislocation patterns and dislocation nucleation in face-centered-cubic and body-centered-cubic bicrystal interfaces. Int J Plast 53:40–55. https://doi.org/10.1016/j.ijplas.2013.07.002

Wang J, Zhou Q, Shao S, Misra A (2017) Strength and plasticity of nanolaminated materials. Mater Res Lett 5:1–19. https://doi.org/10.1080/21663831.2016.1225321.

Was GS, Foecke T (1996) Deformation and fracture in microlaminates. Thin Solid Films 286:1–31. https://doi.org/10.1016/S0040-6090(96)08905-5

Werner E, Prantl W (1990) Slip transfer across grain and phase boundaries. Acta Metall Mater 38:533–537. https://doi.org/10.1016/0956-7151(90)90159-E

Würschum R, Herth S, Brossmann U (2003) Diffusion in nanocrystalline metals and alloys, a status report. Adv Eng Mater 5:365–372. https://doi.org/10.1002/adem.200310079

Xie Y, Han L, An Q, Zheng L, Luo S (2009) Release melting of shock-loaded single crystal Cu. J Appl Phys 105:066103. https://doi.org/10.1063/1.3099597

Yadav SKK, Shao S, Wang J, Liu X-YX-Y (2015) Structural modifications due to interface chemistry at metal-nitride interfaces. Sci Rep 5:17380. https://doi.org/10.1038/srep17380

Yamakov V, Wolf D, Phillpot SR, Gleiter H (2002) Deformation twinning in nanocrystalline Al by molecular-dynamics simulation. Acta Mater 50:5005–5020. https://doi.org/10.1016/S1359-6454(02)00318-X

Yamakov V, Wolf D, Phillpot SR, Mukherjee AK, Gleiter H (2003) Deformation mechanism crossover and mechanical behaviour in nanocrystalline materials. Philos Mag Lett 83:385–393. https://doi.org/10.1080/09500830031000120891

Zbib HM, Diaz de la Rubia T (2002) A multiscale model of plasticity. Int J Plast 18:1133–1163. https://doi.org/10.1016/S0749-6419(01)00044-4

Zbib HM, Rhee M, Hirth JP (1998) On plastic deformation and the dynamics of 3D dislocations. Int J Mech Sci 40:113–127. https://doi.org/10.1016/S0020-7403(97)00043-X

Zhang RF, Wang J, Beyerlein IJ, Germann TC (2011) Dislocation nucleation mechanisms from fcc/bcc incoherent interfaces. Scr Mater 65:1022–1025. https://doi.org/10.1016/j.scriptamat.2011.09.008

Zhang RF, Wang J, Beyerlein IJ, Misra A, Germann TC (2012) Atomic-scale study of nucleation of dislocations from fcc–bcc interfaces. Acta Mater 60:2855–2865. https://doi.org/10.1016/j.actamat.2012.01.050

Zhang RF, Germann TC, Liu XY, Wang J, Beyerlein IJ (2014) Layer size effect on the shock compression behavior of fcc-bcc nanolaminates. Acta Mater 79:74–83. https://doi.org/10.1016/j.actamat.2014.07.016

Zheng S, Shao S, Zhang J, Wang Y, Demkowicz MJ, Beyerlein IJ, Mara NA (2015) Adhesion of voids to bimetal interfaces with non-uniform energies. Sci Rep 5:15428. https://doi.org/10.1038/srep15428

Zhu T, Li JJ, Samanta A, Kim HG, Suresh S (2007) Interfacial plasticity governs strain rate sensitivity and ductility in nanostructured metals. Proc Natl Acad Sci 104:3031–3036. https://doi.org/10.1073/pnas.0611097104

Zhu XY, Liu XJ, Zong RL, Zeng F, Pan F (2010) Microstructure and mechanical properties of nanoscale Cu/Ni multilayers. Mater Sci Eng A 527:1243–1248. https://doi.org/10.1016/j.msea.2009.09.058

# Advances in Discrete Dislocation Dynamics Simulations

# 50

## Richard LeSar and Laurent Capolungo

## Contents

### Abstract

Over the past few decades, discrete dislocation dynamics, a modeling framework allowing for the simulation of the collective motion and interactions of dislocations in crystalline media, has been the subject of intense development worldwide. In recent years, a series of novel numerical algorithms, chemomechanical frameworks, and applications have been proposed. These advances have taken the field closer to enabling predictions of the mechanical response

R. LeSar (✉)
Department of Materials Science and Engineering, Iowa State University, Ames, IA, USA
e-mail: lesar@iastate.edu

L. Capolungo
Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: laurent@lanl.gov

of engineering polycrystals, e.g., textured crystalline aggregates with impurities. Further, interesting pathways have been proposed to bridge discrete dislocation dynamics simulations with harmonic transition state theory, thereby delineating potential routes for performing coarse graining from the viewpoint of thermodynamics. This chapter summarizes some of the important recent contributions in the field of discrete dislocation dynamics.

# 1    Introduction

Plastic deformation in crystalline metals is generally a consequence of the collective motion of large numbers of curvilinear defects called dislocations. The mobility of dislocations gives rise to plastic flow at relatively low stress levels compared to the theoretical strength. In a typical metal, dislocation densities, $\rho$, range from $10^{10}$ to $10^{15}$/m$^2$, i.e., $10^{10}$–$10^{15}$ m of dislocations per cubic meter of material, values that typically increase rapidly under applied stress (or strain). Further, dislocations form organized structures such as pile-ups, walls, and cells (typically referred to as dislocation substructures). The topological constraints arising from the crystallography greatly add to the complexity of describing dislocation evolution and dynamics. More information about dislocations can be found in many places, for example, in Bacon (1992), Hull and Bacon (2001), and Anderson et al. (2017).

The length scale of dislocation substructures is on the order of many microns, well beyond what can be studied with atomistic-based methods like molecular dynamics. To capture the complexity of correlated and collective dislocation motion and interactions requires a *mesoscale* approach, in which both the length and time scales are extended – in comparison to atomistic simulations – by using dislocations as the simulated entities. In discrete dislocation dynamics simulations (DDD), the elastic forces on dislocations are calculated, and the equations of motion are solved, yielding a quantification of plastic strain as a function of an applied load. While many interesting results have been found using simple two-dimensional models (e.g., Balint et al. 2008; Quek et al. 2014; Zheng et al. 2016), the focus of this paper will be on fully three-dimensional DDD simulations.

The purpose of this chapter is to discuss recent advances in DDD simulations, so in Sect. 2 we give only a brief overview of the basics of dislocation simulations while pointing out some of the limitations in the methods as usually employed. In Sect. 3, recent advances in algorithms will be discussed, including methods to extend the time scales (Sect. 3.1) and recent Green function methods, used to calculate the stresses arising from the dislocations, that are solved using fast Fourier transforms (FFT) in Sect. 3.2. Recent advances in applications of DDD are discussed in Sect. 4, with a specific focus on polycrystalline materials in Sect. 4.1 and the coupling of DDD with chemistry (Sect. 4.2). Improvements in models for sub-scale physics are discussed in Sect. 5, focusing on models for transmission across boundaries (Sect. 5.1) and kinetic models for activated processes (Sect. 5.2). A summary of the state of the field and a discussion of what new capabilities are needed are captured in Sect. 6.

## 2       Basics of Dislocation Dynamics Simulations

DDD simulations all follow the same general procedure, as shown schematically in Fig. 1, which indicates the types of calculations needed and the required information at each step. Here we present the basics of what happens at each time step, with more details available from many sources, such as (Fivel 2008a, b; Sills et al. 2016b; LeSar 2014), recent books (Bulatov et al. 2006; Kubin 2013), and another chapter in this Handbook ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" by Sills and Aubry.

   Dislocations are generally represented as a series of nodes connected by straight lines (Kubin and Canova 1992; Devincre and Kubin 1997; Zbib et al. 2000) or spline-fitted curves (Ghoniem et al. 2000; Wang et al. 2006), as shown in Fig. 1a. As the dislocation network evolves, the nodes must be adjusted (referred to as remeshing). In Fig. 1a, the red segment (connecting nodes 3 and 4) is a *junction* between the dislocation connecting nodes {12345} and the dislocation connecting nodes {7346}. More details on junctions in dislocation dynamics simulations are described in another chapter in this Handbook ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" by Sills and Aubry.
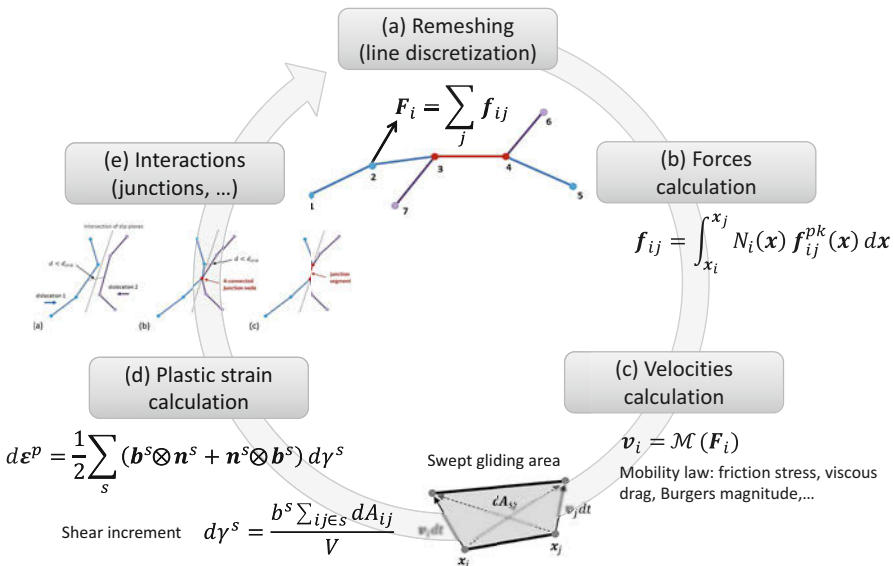


**Fig. 1** Diagram of the main stages composing the basic cycle performed at each time step in DDD simulations. (a) Dynamic discretization of dislocation lines: the topology of the dislocation network evolves at each time step requiring discretization to be performed dynamically. (b) Forces calculation: the stress driving dislocation motion is evaluated from the spatial configuration of the dislocation network. (c) Velocities calculation: once forces on dislocations have been computed, the motion of dislocation lines can be calculated through the mobility law. (d) Plastic strain calculation: the areas swept by dislocation motion allow for the determination of the plastic activity. (e) Interactions: collisions between dislocations during glide are treated via topological rules. The DDD cycle (a) to (e) is repeated until simulation is completed

Once the dislocation network is discretized, the forces on the dislocation segments must be calculated. These forces are determined by calculating the stress at each segment and then applying the usual Peach-Koehler force expression (described in Anderson et al. 2017). For each dislocation node $i$, the nodal force $\mathbf{F}_i$ is the sum of the contributions by each segment $ij$ connected to node $i$ (Fig. 1a):

$$\mathbf{F}_i = \sum_j \mathbf{f}_{ij} , \tag{1}$$

in which $\mathbf{f}_{ij}$ is the force on segment $ij$ acting at node $i$ and is found by integration of the effective force acting along the dislocation segment, as shown in Fig. 1b. More complete details of the calculation of forces on a node are given in Bulatov and Cai (2006).

The stresses from the dislocations are generally calculated by adding up analytical expressions that were derived assuming isotropic elasticity, e.g., as a sum over individual straight segments (e.g., as presented in Anderson et al. 2017) or, for curved segments, as an integral over the dislocation line (de Wit 1960; Ghoniem and Sun 1999). In another chapter of this Handbook ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" by Sills and Aubry, recent advances for calculating stresses based on anisotropic elasticity are discussed.

A major challenge is that the stresses from a dislocation are long ranged and special methods are needed to sum those interactions efficiently. A particularly effective method is the fast multiple method, which has been used in both two-dimensional simulations (Wang and LeSar 1995) and, more recently, in three dimensions (Arsenlis et al. 2007). Another approach, called the box method (Verdier et al. 1998), is less accurate but easier to implement and has less computational overhead and thus likely to be more computationally efficient for small systems. Given that the stress calculation represents the major computational burden, much effort has been made to increase its efficiency through the use of highly parallel computational methods, as in, for example, Arsenlis et al. (2007) and Wang et al. (2006). An alternative approach, based on a direct fast Fourier transform (FFT)-based integration of the Green function to calculate the stress, is described in Sect. 3.2.

The equations of motion for the dislocations are then solved for a given time step. At high strain rates, the full damped dynamics equations of motion are required (Wang et al. 2007) for each node:

$$\frac{d\left(m_i^\star \mathbf{v}_i\right)}{dt} = \mathbf{F}_i - \gamma \mathbf{v}_i , \tag{2}$$

in which, for each node $i$, $m_i^\star$ is the effective mass (which depends on the velocity as described in Hirth et al. 1998), $\mathbf{v}_i$ is the velocity, $\mathbf{F}_i$ is the force on the node from Fig. 1a, and $\gamma$ is the damping coefficient. For strain rates in the range of $10^4 \, \mathrm{s}^{-1}$ and below, it is usually sufficient to assume that the time to reach terminal velocity is small relative to the time step used to solve Eq. (2) (Wang et al. 2007) and equations assuming overdamped dynamics can be employed, with the velocity of the ith node given by

$$\mathbf{v}_i = \mathbf{M} \cdot \mathbf{F}_i, \tag{3}$$

in which $\mathbf{M}$ is the mobility tensor. Mobility laws are discussed in ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" by Sills and Aubry.

The simplest approach to solve for the new nodal position ($\mathbf{R}_i$) is to use a linear, forward, Euler integration based on Eq. (3) and a time step $\delta t$:

$$\mathbf{R}_i(t + \delta t) = \mathbf{R}_i(t) + \mathbf{v}_i(t)\delta t. \tag{4}$$

The choice of time step in Eq. (4) is critical, for if the nodal displacement ($\Delta\mathbf{R}_i(t) = \mathbf{v}_i(t)\delta t$) approaches the size of the segments to which node $i$ is connected, instabilities can arise. One way to minimize instabilities is to use a dynamic time step, which can be done in a number of ways. A particularly simple approach is to prescribe the maximum distance of a nodal displacement, $\delta R_{max}$, calculating the absolute value of the maximum force on a node $|F|_{max}$ at each time step and setting the time step by $\delta t = \delta R_{max}/|F|_{max}$. $\delta R_{max}$ can be varied to achieve stable integration of the nodal positions; however, the time steps can be very small (on the order of $10^{-13}$ s or less) for situations in which dislocations are close together (Wang et al. 2007).

A major issue with the simple integration scheme just described is that there is no way to set an accuracy of the time integration. One approach to improve the time integration and allow for longer time steps is by the use of "predictor-corrector" methods. Predictor-corrector methods use an iterative procedure to find the change in nodal positions. For example, the Heun's method (used in the commonly used DDD program ParaDiS) is a second-order accurate explicit method that uses a forward predictor (Sills et al. 2016a):

$$\mathbf{R}^0(t + \delta t) = \mathbf{R}^0(t) + \delta t\, \mathbf{v}(t), \tag{5}$$

coupled with a trapezoidal method corrector,

$$\mathbf{R}^{j+1}(t + \delta t) = \mathbf{R}^0(t) + \frac{\delta t}{2}\left(\mathbf{v}(t) + \mathbf{v}^j(t)\right), \tag{6}$$

in which $\mathbf{R}$ is the nodal position, $\mathbf{v}$ is the nodal velocity, and the superscript $j$ indicates the iteration number. The error is estimated as

$$E_{iter} = ||\mathbf{R}^{j+1}(t + \delta t) - \mathbf{R}^j(t + \delta t)||, \tag{7}$$

in which $||\cdot||$ is the $L_2$ norm. An iteration proceeds until $E_{iter}$ is less than a prescribed parameter. In Sills et al. (2016a), other, more complicated and more accurate, time integration algorithms are discussed. The time step is dynamically chosen to control the absolute error at each time step, with an error tolerance set by the user. Details are provided in Sills and Cai (2014).

Sills et al. (2016a) described the effectiveness of time integration schemes in simulations of strain hardening in samples with dislocation densities in the range of 0.5 to $3 \times 10^{12}$ m$^{-2}$. When using the Heun algorithm, the authors found that accurate integration of the equations of motion required a range in time steps between $10^{-14}$

and $10^{-11}$ s, with an average time step of $6.5 \times 10^{-13}$ s. Such small steps limit the utility of DDD simulations to relatively small total times, which in turn limits the use of DDD to achieve high levels of strain when using relevant strain rates. In Sect. 3.1, we will discuss recent progress on advanced algorithms for time integration.

In part (d) of Fig. 1 are shown the two stages for calculating the net plastic strain after having determined the dislocation motion. First the area swept by the dislocation is calculated, yielding the amount of shear on its slip plane, $d\gamma^s$ in Fig. 1. The total plastic strain increment $d\varepsilon^p$ is the projection of the shear summed over all slip systems $s$.

The final step of a simulation involves the inclusion of various models that describe sub-scale physics, i.e., physics that is inherently atomistic in nature, which includes the formation of junctions, annihilation, and dissociation. The computational challenges include the detection of possible events and the topological procedures needed to ensure that the correct physics is being modeled. Some aspects of these models are discussed elsewhere in this Handbook ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" by Sills and Aubry.

All three-dimensional discrete dislocation dynamics calculations involve these basic steps. In the next section, we discuss new algorithmic developments that extend the capabilities of DDD simulations to longer times and more complex microstructures.

## 3    Recent Advances in Algorithms

The key computational challenges of discrete dislocation dynamics as usually employed are the time it takes to calculate the stresses, the use of equations for the stress based on isotropic elasticity, and the very small time steps needed to resolve motion when dislocations are close together (discussed in Sect. 3.1). As noted above, much effort has been spent developing ways to reduce the time needed for calculating the stresses by using ways to speed up the calculation of long-range stresses, the use of parallel computation, and, more recently, the use of GPUs.

As discussed elsewhere in this Handbook ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" by Sills and Aubry, the inclusion of anisotropic elasticity has been addressed by many researchers, with a number of methods having been developed. Most of these methods, however, are computationally intensive (over 200 times calculations based on isotropic elasticity) and thus have been rarely used (Yin et al. 2010). Recent work by Aubry and coworkers (Aubry and Arsenlis 2013) using a spherical harmonic expansion of the Green function has led to increased efficiencies of the anisotropic calculations. The computational expense of these methods is highly dependent on the anisotropy factor, ranging from a 50% increase for an anisotropy factor of about 2 (aluminum) to a factor of 60 increase for an anisotropy factor of 7.4 (iron). In Sect. 3.2, we will discuss a new approach that reduces the excess computational burden even further.

## 3.1   Extending Time Scale

The question of finding efficient time integration schemes is recognized as a key issue for discrete dislocation dynamics, as discussed in a series of recent papers (Gardner et al. 2015; Sills and Cai 2014; Sills et al. 2016a). In this section, we focus on recent work that extends the time scales to enable simulations of complex phenomena that require long times, for example, models of work hardening that achieve strains relevant to experiment.

As noted in Sect. 2, the time steps needed for accurate solutions to the equations of motion in discrete dislocation dynamics calculations are typically small enough to limit simulations to small plastic strains. The small time steps are the result of dislocations that are sufficiently close together so that the net forces between them are quite high and nonlinear. The motion of dislocations that are not close to other dislocations will generally have slowly varying interactions and can be accurately calculated with much larger time steps. The disparity of time steps suggests the need for a method that employs time step subcycling for those dislocations that have nearby dislocation and require small time steps.

Cai and coworkers (Sills et al. 2016a) present a specific approach for efficient time integration that is based on the use of the `ParaDiS` code. `ParaDiS` represents dislocations as a series of nodes connected by segments and employs the O(N) fast multipole method (FMM) to calculate the long-ranged stress fields. In the FMM, the net stress at a point is determined by direct sums of the pairwise stresses for nearby dislocations and a hierarchical multipole expansion for more distant dislocations. `ParaDiS` employs the Heun's method from Eqs. (5), (6), and (7) to do the time integration. In Sills et al. (2016a), they also examine the use of more advanced time integration schemes, but we will restrict the discussion to results with the Heun method.

The method of Sills et al. (2016a) is a force-based subcycling. The set of forces on the segments are divided into groups based on the minimum distance between the segments and their neighboring segments, as shown schematically in Fig. 2a. The forces between all segments that lie between zero and $r_2$ are assigned to group 2, forces for segments between $r_2$ and $r_3$ are in group 3, forces between segments with distances between $r_3$ and $r_4$ are in group 4, and all other forces are assigned to group 0, the global group.

Based on velocities of the segments (Eq. (3)), segments in the global group (0) are evolved using a standard time evolution algorithm (e.g., Heun), and the global time step is determined. For segments in other groups, a sequential splitting method (Geiser 2009) is employed, continuing until the time integration converges. Additional procedural and details are described in Sills et al. (2016a).

The subcycling time integration approach was tested (by comparing the calculations based on standard time integration Sills et al. 2016a) for the same simulations of strain hardening in samples with dislocation densities in the range of 0.5 to $3 \times 10^{12} \, \mathrm{m}^{-2}$ described above. For these studies, the groups were created using $r_{g2} = 100b$, $r_{g3} = 600b$, and $r_{g4} = 1600b$, where $b$ is the Burgers vector. As noted above, with the Heun algorithm and no subcycling, the average time step

**Fig. 2** (**a**) Schematic view of how interaction forces are assigned to subcycle groups based on the minimum distance between segments. The dashed ovals represent regions of constant minimum distance from the segment centers at distances of $r_{g2}$, $r_{g3}$, and $r_{g4}$. In their approach, $r_{g1} = 0$ so the segment's self-interaction is assigned to group 1. Colors and numbers indicate the group. (Adapted from Sills et al. 2016a)
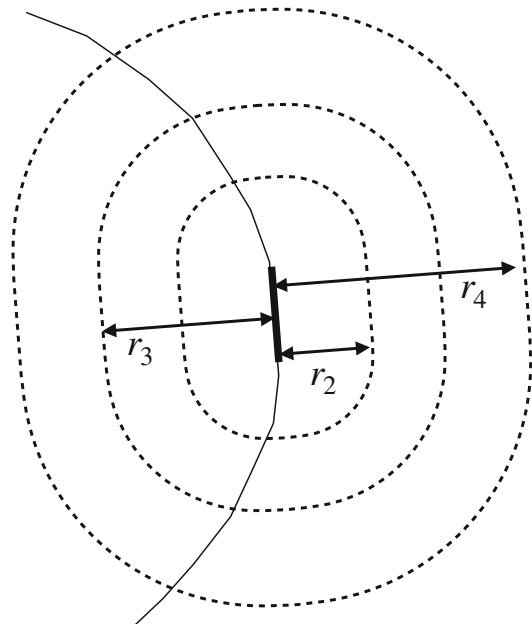


was $6.6 \times 10^{-13}$ s. For the same system with time subcycling, the average global time step (with subcycling) was extended to $1.8 \times 10^{-10}$ s (Sills et al. 2016a). The consequence of the longer time step is clear: for the same clock time, the Heun algorithm without subcycling could calculate a total strain of about 0.13%, while with subcycling, the net strain was about 0.6%. Overall, the authors found that the global time step was approximately a factor of 100 larger than in simulations with no subcycling.

These advances in subcycled time integration offer new possibilities for achieving higher calculated strains in discrete dislocation dynamics simulations, which enhances the ability to study such important phenomena as strain hardening. It remains unproven, however, how much these methods will affect simulations at the high dislocation densities common in well-worked metals ($10^{15}$ m$^{-2}$). In Sills et al. (2016a), the reported dislocation densities ($10^{12}$ m$^{-2}$) were considerably smaller, and in dense networks, the fraction of dislocations in the "global group" may be negligible and the benefits of subcycling the time integration reduced.

## 3.2   Fast-Fourier Transform Methods

Stresses in a lattice can also be calculated by identifying the *eigenstrains* associated with the presence of defects. From Mura (1987), an eigenstrain is a nonelastic strain, arising from several potential effects, such as thermal expansion misfits, phase transformations, plastic strains, or misfit strains. Eigenstresses are the self-equilibrated internal stresses caused by the eigenstrains. Here we show how to

calculate the stresses from a set of dislocations such as to simulate their dynamic evolution (Bertin et al. 2015; Graham et al. 2016).

We decompose the total strain into elastic strains and eigenstrains, denoted by $\varepsilon_{ij}^e$ and $\varepsilon_{ij}^*$, respectively, i.e.:

$$\varepsilon_{ij}(\mathbf{r}) = \varepsilon_{ij}^e(\mathbf{r}) + \varepsilon_{ij}^*(\mathbf{r}) . \tag{8}$$

Let us denote the displacement field within the system with $u_i$. The total strain is given in terms of the distortions, which are the displacement gradients $u_{i,j} = \partial u_i / \partial x_j$, in which a subscript comma indicates a derivative. For example, the elastic strain is given by the usual expression:

$$\varepsilon_{ij}^e = \frac{1}{2} \left( u_{i,j}^e + u_{j,i}^e \right) \tag{9}$$

and the total distortion can be written as

$$u_{i,j}(\mathbf{x}) = u_{i,j}^e(\mathbf{r}) + u_{i,j}^*(\mathbf{r}) . \tag{10}$$

Starting with the equation for mechanical equilibrium, $\sigma_{ij,j} = 0$, it is straightforward to show that the elastic distortions are given by

$$u_{k,l}^e = u_{k,l} - u_{k,l}^* = -G_{ki,l} * (c_{ijmn} u_{m,nj}^*) - u_{k,l}^* , \tag{11}$$

in which $G_{km}$ is the elastic Green tensor (described in detail elsewhere de Wit 1960) and $g * f = \int g(\mathbf{x} - \mathbf{x}') f(\mathbf{x}') d\mathbf{x}'$ is the convolution integral. The Fourier transform of Eq. (11) is

$$\hat{u}_{k,l}^e(\mathbf{k}) = \hat{G}_{ki} c_{ijmn} k_j k_l \hat{u}_{m,n}^* - \hat{u}_{k,l}^* , \tag{12}$$

in which a Fourier transform of a function $f$ is denoted by $\hat{f}$, $k_i$ is a component of a wave vector, and $c_{ijkl}$ are elastic constants. $\hat{G}_{ki}$ is the Fourier transform of the elastic Green tensor, whose inverse is given by de Wit (1960)

$$\hat{G}_{ik}^{-1} = c_{ijkl} k_j k_l . \tag{13}$$

Given the Fourier transform of the distortion associated with the eigenstrain, $\hat{u}_{m,n}^*$, we have an algebraic expression for the elastic distortion, $\hat{u}_{k,l}^e(\mathbf{k})$, at each point in $k$-space.

A fast Fourier transform (FFT) method is used to calculate $\hat{u}_{i,j}^*(\mathbf{k})$ at each point of the FFT grid ($\mathbf{x}$). Equation (12) is used to evaluate $\hat{u}_{k,l}^e(\mathbf{k})$, from which $\hat{\varepsilon}_{kl}^e(\mathbf{k})$ is determined. We take the inverse transform to find $\varepsilon_{kl}^e(\mathbf{x})$ using Eq. (9). The stress at each FFT grid point is

$$\sigma_{ij}(\mathbf{x}) = c_{ijkl}\,\varepsilon_{kl}^e(\mathbf{x})\,. \tag{14}$$

Since there are no constraints on the values of $c_{ijkl}$, there is no extra penalty for using the full anisotropic elastic constants in computing the stress and strain fields.

These equations are applicable for any eigenstrain or combination of eigenstrains. The eigenstrain associated with a dislocation is the co-called *plastic distortion tensor*, which is denoted by $\beta_{ij}^p$ and is the distortion associated with the plastic strain:

$$\beta_{ij}^p = u_{i,j}^p\,. \tag{15}$$

$\beta_{ij}^p$ replaces $u_{i,j}^*$ in Eq. (12).

The plastic distortion tensor is a direct measure of the incompatible slip caused by a dislocation. Following Mura (1987), the distortion tensor for a straight edge dislocation located at $(x_1, x_2) = (0, 0)$ with Burgers vector along $\hat{x}_1$ and line direction along $\hat{x}_3$ has only the 21 component that is non-zero. The distortion tensor is

$$\beta_{21}^p = b\,\delta(x_2)\Theta(x_1)\,, \tag{16}$$

where $\Theta(x)$ is the Heaviside step function defined by $\Theta(x) = 1$ for $x < 0$ and 0 for $x > 0$ and $b$ is the magnitude of the Burgers vector. $\delta(x_2)$ is a Dirac delta function that restricts the slip to the plane defined by $x_2 = 0$. Thus, $\beta_{21}^p$ describes the slip (with magnitude $b$) introduced in the upper half plane by the dislocation. Similarly, for a circular loop in the $x_1 x_2$ plane with $\hat{b} = \hat{x}_1$, there is slip inside the loop but not outside the loop – the only non-zero component of the plastic distortion tensor is $\beta_{31}^p$. In all cases, a dislocation line can be identified by a discontinuous jump in $\beta_{ij}^p$.

Recent papers Bertin et al. (2015) and Graham et al. (2016) have implemented this general procedure for calculating the stresses needed for discrete dislocation dynamics simulations. Detailed comparisons in each paper show that the stresses arising from distributions of dislocations calculated with the FFT method agree with those determined from standard analytical expressions (Anderson et al. 2017) for isotropic elasticity. They also show, however, a standard problem that arises when using Fourier transform methods with discontinuous functions (e.g., $\beta^p$) – spurious oscillations and spiking in calculated quantities. Such oscillations, often referred to as Gibbs oscillations, occur frequently in signal processing and computer graphics applications, and many work-arounds have been proposed, mostly focused on smoothing the discontinuity over the Fourier grid. In Graham et al. (2016), for example, an antialiasing method from computer graphics coupled with a low-pass filter was shown to remove the oscillations with minimal effect on the final calculated stresses. Bertin et al. (2015) used an approach in which the dislocation density was spread in a three-dimensional triangular distribution over a $3 \times 3 \times 3$ set of voxels, again with minimal effect on the calculated stresses. This numerical

spreading is analogous to the spread of a dislocation near its core, though with a core width that is related to the mesh size. More recently, replacing the Fourier transform of the derivatives with a transform of discrete derivatives has also been successfully employed to attenuate the Gibbs oscillations (Graham et al. 2019).

Based on the above FFT stress calculations, two approaches for using those stresses in discrete dislocation dynamics calculations were proposed. In Graham et al. (2016), the size of the FFT grid was set such that only one dislocation segment was allowed in each voxel (the three-dimensional volume between grid points). This choice limits the simulations to relatively small sizes and densities of dislocations. The other approach assumes that many dislocations can lie in each voxel (Bertin et al. 2015), which has some complications but enables simulations of more relevant dislocation distributions, as will be summarized next.

The main components of the implementation of the FFT stress calculation within a discrete dislocation dynamics framework in Bertin et al. (2015) are the following: the distortions in Eq. (12) are replaced by the strain tensor $\varepsilon_{kl}$ (which can be shown to be equivalent), the plastic strain change at each time step is calculated by measuring the area swept by the dislocation movement, and by coupling the FFT-based approach to the discrete continuous model (DCM) (Lemarchand et al. 2001; Vattré et al. 2014a). By using the DCM, the FFT approach takes advantage of the high computational efficiency of the FFT algorithm for calculating the long-range forces while allowing for a discrete representation of local dislocation interactions. The computational time associated with the new DDD-FFT approach is significantly lower than that of other DDD approaches when large number of dislocation segments are involved (Bertin et al. 2015). Moreover, the use of anisotropic elasticity comes at a similar computational cost to that of an isotropic simulation. In Sect. 4.1, this approach forms the basis of a new approach for polycrystal plasticity based on discrete dislocation dynamics.

The promise of the FFT approach is that it scales essentially linearly with the number of dislocation segments and, with the use of GPUs to evaluate the stresses, is computationally very fast, such that the stress calculation no longer dominates the time to perform a calculation. This approach opens up the possibility of extending DDD to the high dislocation densities present in such important problems as well-worked metals and fatigue.

## 4   Recent Advances in Applications

Advances in algorithms have enabled the use of DDD for a number of new applications. One such example is the use of new time integration schemes to extend calculations of strain hardening to longer times and, thus, larger strains, as discussed in Sect. 3.1. Here we focus on two other examples, the first being the ability, based on the FFT method described above, to do simulations of the deformation of polycrystals with arbitrary grain sizes, shapes, and orientations. The second example is the linkage between dislocation dynamics and chemistry, an important advance for understanding the effects of chemical constituents on deformation.

## 4.1    Polycrystal Plasticity

A number of studies of dislocation dynamics in polycrystalline materials have been made using two- and three-dimensional models. For this chapter, we restrict the discussion to simulations in three dimensions. We also focus our discussion of calculations in three dimensions on those in which the deformation in all grains in a polycrystalline sample is calculated, rather than in just a single, representative, grain.

All of the studies to date using DDD to study deformation in polycrystalline have been on model systems that do not directly reflect the grain sizes, shapes, and orientations found in general polycrystalline materials. (For more details of prior applications on DDD in polycrystals, see Sect. 5.1.) For example, one study coupled finite element polycrystal plasticity with discrete dislocation dynamics using a system of nine uniform cuboidal grains located in a plane with periodic boundary conditions (Siška et al. 2009). Each grain was oriented along a few specific crystal directions. Others used the same system of nine grains to study the effects of grain size and aspect ratio on the plasticity of free-standing polycrystalline films; however in this work all the grains had the same orientation and the boundaries between the grains served only as stress-controlled barriers to transmission (Zhou and LeSar 2012). Recent simulations were done on pseudo bi-crystals with varying grain disorientation (Prasad Reddy et al. 2013) while other studies examined the role of twins in magnesium (Fan et al. 2015a, b, 2016) and laths in iron (Chaussidon et al. 2008). The most complete model of polycrystalline plasticity at the discrete dislocation dynamics level was in a recent calculation in which the system was represented with a set of equal-sized truncated-octahedral grains with random orientations (de Sansal et al. 2010). While the grain shapes were fixed (and uniform), the effects of the grain size and orientation on the deformation were studied. Overall, however, all of these studies approximated the grain distributions in some way.

Modeling the deformation in polycrystalline systems can be accomplished by coupling polycrystal plasticity calculations, in which the constraints placed on grain evolution by their neighboring grains are directly modeled in the calculation, with discrete dislocation dynamics simulations. The approach described here uses an FFT-based polycrystal plasticity model linked to the FFT-based DDD model described in Sect. 3.2.

Building on previous work (Moulinec and Suquet 1998; Michel et al. 1999), Lebensohn (2001) and Lebensohn et al. (2012) introduced a description of crystal plasticity based on the same basic fast Fourier transform (FFT) algorithm described above but modified to solve an eigenstrain description of the elasticity of heterogeneous media (Mura 1987). The FFT approach solves for the equilibrium stress and strains under the constraint of strain compatibility for systems with periodic microstructures. The plasticity of the polycrystal is described by a parametrized constitutive model, for example, an elasto-viscoplastic model in Lebensohn et al. (2012). The computational time can be several orders of magnitude shorter for the FFT than in the traditional finite element methods (FEM). The results of the FFT approach have been compared in detail with similar calculations based on FEM for a

periodic system of face-centered cubic grains (Liu et al. 2010). The final textures and strain distributions calculated with the two methods show some differences which are still under investigation, in which the FFT approach appears to be less dependent on the resolution of the grids than FEM and, indeed, predicts strain gradients very well at quite low resolution.

The challenge with polycrystalline materials is that the elastic tensor of a grain in the system frame is different depending on the orientation of the grain. To treat such heterogeneous elastic media, Lebensohn, following Moulinec and Suquet (1998) and Michel et al. (1999), introduced a reference stiffness tensor $C_{ijkl}^o$, with details given in Lebensohn et al. (2012). It is relatively straightforward to show that in this approach, the equations for the stress given in Eqs. (12) and (13) are replaced by

$$
\sigma_{ij}^{tot}(\mathbf{x}) = C_{ijkl}^o \varepsilon_{kl}(\mathbf{x}) + \tau_{ij}(\mathbf{x})
$$
$$
\tau_{ij}(\mathbf{x}) = \delta C_{ijkl}(\mathbf{x}) \varepsilon_{kl}(\mathbf{x}) - C_{ijkl}(\mathbf{x}) \varepsilon_{kl}^p(\mathbf{x}) \tag{17}
$$
$$
\delta C_{ijkl}(\mathbf{x}) = C_{ijkl}(\mathbf{x}) - C_{ijkl}^o ,
$$

in which $\mathbf{x}$ is an FFT grid point, $C_{ijkl}(\mathbf{x})$ is the stiffness tensor in the grain associated with $\mathbf{x}$, and $\delta C_{ijkl}(\mathbf{x})$ is a fourth-order tensor quantifying the deviation in the elastic properties associated with each material point $\mathbf{x}$ from that of the reference medium. Note that the solution to the equations in Eq. (17) must be solved in an iterative fashion (Lebensohn et al. 2012).

The FFT-based dislocation dynamics approach (Bertin et al. 2015) described in Sect. 3.2 has recently been implemented (Graham et al. 2019) within the polycrystal plasticity method of Lebensohn et al. (2012). A three-dimensional polycrystalline sample with random grain size shape can either be created through a Voronoi construction or from experimentally determined microstructures. The orientation of the grains is either set randomly or from experiment. An appropriate FFT grid is overlaid on the 3D sample, and periodic boundary conditions are employed. The grain in which each FFT grid point is located is determined, and the stiffness tensor in the system frame is determined from the grain orientation. Typical simulations are done by seeding an initial random distribution of dislocation loops on the slip systems within each grain. For example, in the example discussed below, dislocation loops were placed randomly on the 12 slip systems of a face-centered cubic system within each grain, and the initial plastic strain determined by a method equivalent to evaluating the plastic distortion tensor of Eq. (15). The DDD method described in Sect. 3.2 (Bertin et al. 2015) was then employed, tracking the total deformation of the system as well as the deformation in each grain. The outcome is a polycrystal plasticity simulation package in which a constitutive model of plasticity is replaced by the discrete dislocation dynamics approach described in Bertin et al. (2015).

An example calculation is shown in Fig. 3, in which a system of six grains of nickel with random size, shape, and orientation is included in a periodic cube 1.5 μm per side. The grains were created with a Voronoi construction and mapped onto the FFT grid. A time step of $10^{-13}$ s was used. After an initial relaxation sequence

**Fig. 3** Example calculation of a system of six grains of nickel with random size, shape, and orientation is included in a periodic cube 1.5 μm per side. Shown is the dislocation distribution after an initial relaxation sequence of 1000 time steps, followed by an additional 5000 time steps under stress. The dislocations are shown in the same color as the grain in which they are located
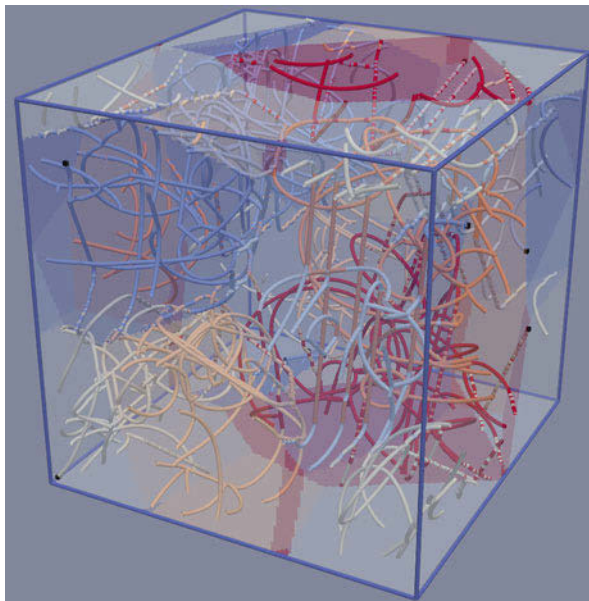


of 1000 time steps, an additional 5000 time steps were performed. At each time step, the iterative solution described in Eq. (17) was used to determine the stress at each FFT grid point. Figure 3 shows the final configuration of that run, with the dislocations shown in the same color as the grain in which they are located. In this model, the interactions between dislocations and the interfaces are approximated with a simple transmission model.

The combined FFT-based polycrystal plasticity/discrete dislocation dynamics (PP/DDD) approach is, to our knowledge, the only such method available for three-dimensional polycrystal plasticity of grains with arbitrary size, shape, and orientation based on the DD simulation of the motion of individual dislocations. Given the linear scaling of the FFT with the number of dislocation segments, there are many potential applications of the PP/DDD approach to important problems in the deformation of materials. There are, however, numerous advances in physics that are needed for accurate polycrystal plasticity calculations, the most important of which are the interactions of dislocations with grain boundaries, which are discussed in Sect. 5.1.

## 4.2 Coupling Dislocation Dynamics with Chemistry

With the intent to develop concurrent multiscale/multiphysics modeling schemes predicting microstructure evolution due to dislocation motion and interaction as well as point and defect cluster rearrangements, a series of studies have focused on coupling mean-field rate theory (MFRT) (Braislford and Bullough 1981), or one

of its derivatives such as cluster dynamics, and DDD. In the former, one establishes continuity equations for each chemical species considered. These essentially translate the effects of point defects and defect cluster generation, migration, and interactions (i.e., annihilation, absorption, emission) on the local concentration of the species of interest. As such, the treatment of dislocation/vacancy interactions has seen a significant gain in interest over the past decade. Fundamentally, this problem holds the key to developing predictive laws of the contribution of dislocation climb to plasticity. In a relatively large series of experimental studies focused on the creep response of metals, the discernment of the contribution of cross-slip and climb to the overall creep rate has been problematic (Kombaiah and Murty 2015; Morrow et al. 2013). More generally and as discussed in Keralavarma et al. (2012) and Keralavarma and Benzerga (2015), the hope is to be able to perform a virtual mapping of the prominence of specific dissipative processes (i.e., glide, cross-slip, climb), similar to that proposed by Frost and Ashby (1982), by means of a hybrid DDD/MFRT approach. Overall, two different types of challenges have to be addressed. First, from the numerical standpoint, the question of time integration is not trivial. The relative rates associated with both processes depend on stress and temperature. Second, one must be able to quantify the two-way coupling between the mechanical and "chemical" problem. Indeed, in the short range, the net current of vacancies toward the dislocation leads to a change in the dislocation geometry (i.e., creation of jogs). However, the long-range stress field inherent to the dislocation and to the point defects necessarily modifies the kinetics of point defect migration. Further, in the very short range, modeling of pipe diffusion requires extraneous considerations to be detailed later in this section. Focusing first on the physical aspects of the problem. Most proposed schemes rely on early analytical developments proposed by Hirth and Lothe summarized in Anderson et al. (2017). The net climb rate of a dislocation segment can be directly related to the vacancy current into the segment. One can relate the climb rate of a unit length segment, $v_{climb}$, to the net vacancy current (i.e., difference between the interstitial and vacancy currents into the segment) into said segment $I$ as follows:

$$v_{climb} = \frac{I.\Omega}{L_{seg}|\mathbf{b} \times \boldsymbol{\xi}|} \tag{18}$$

Here $\Omega$, $\mathbf{b}$, $\boldsymbol{\xi}$, and $L_{seg}$ denote the atomic volume, Burger's vector, the unit tangent vector to the line, and the length of the segment considered, respectively. To quantify the net vacancy current toward the dislocation segment, most DDD-based studies disregard the flux of interstitials toward the dislocation (despite the preferred attraction of those toward dislocations) and focus solely on vacancies. One then must solve the diffusion problem formally posed by

$$\frac{dC}{dt} = -\nabla \cdot \mathbf{J}, \tag{19}$$

in which $C$ denotes the spatially varying vacancy concentration. The dot product between operator Nabla and the vacancy flux, $\mathbf{J}$, denotes the divergence of the flux vector. The latter is formally expressed as

$$\mathbf{J} = -D\left(\nabla C + \beta C \nabla E\right), \tag{20}$$

in which $D$ denotes the vacancy diffusivity and $\beta = 1/kT$ with T the temperature in Kelvin and $k$ the Boltzmann constant. One notes that the flux depends on two distinct contributions. The first term is related to the gradient of the concentration and is purely chemical, while the second involves the interaction energy between the defect and the surrounding microstructure, $E$. It is typically expressed as the double dot product between the local elastic strain and an interaction dipole tensor $d$. With this, one has $E = \varepsilon_{ij}d_{ij}$. Unfortunately, even in the simplest case, an isolated dislocation interacting with a cloud of vacancies with equilibrium concentration at zero stress $C_{eq}$, there is no analytical solution to the aforementioned problem, owing to the non-radial flux induced by the $E$. Thus far, two solutions have been proposed. First, it has been suggested to simply neglect the elastic interactions between the point defect and the dislocation. Such simplification renders the problem entirely radial. The solution procedure then relies on determining the boundary conditions of the problem from which the flux toward the dislocation can be found in the steady state. The local concentration near the dislocation core is given as:

$$C_{core} = C_{eq}e^{\beta P \Delta V} = C_{eq}e^{\dfrac{F_{climb}\beta\Omega}{L_{seg}|\mathbf{b}\times\boldsymbol{\xi}|}}, \tag{21}$$

in which $P$ denotes the hydrostatic pressure within the medium and $\Delta V$ refers to the relaxation volume associated with the vacancy such that $P\Delta V$ denotes the work associated with the creation of a vacancy in the core of dislocation. As shown in the rightmost part of Eq. (21), this work is balanced by the work done by the dislocation as it climbs, in a fashion similar to that proposed by Hirth and Lothe (Anderson et al. 2017). With the steady-state vacancy concentration, the solution to Eq. (19) is as follows:

$$C(r) = C_{core} + (C_{eq} - C_{core})\dfrac{\ln\dfrac{r}{r_{\infty}}}{\ln\dfrac{r}{r_{core}}}. \tag{22}$$

Here $r$, $r_{core}$, and $r_{\infty}$ denote the distance from the dislocation core, the critical core radius, and the distance from the dislocation at which the equilibrium concentration is reached. The latter is often taken as the mean spacing between dislocations. One notes that in all such derivations, the interaction energy between the dislocation and the vacancy is disregarded. In any case, the current toward the dislocation is then simply written as the integral around the core of the dislocation of the vacancy flux. One has

$$I = -2\pi r_{core} L_{seg} \frac{D}{\Omega} \frac{\partial C(r)}{\partial r}\bigg|_{r=r_{core}} \tag{23}$$

This approach was used in a series of studies (Mordehai et al. 2008; Bako et al. 2011; Danas and Deshpande 2013; Gao et al. 2017), both two and three dimensional. Among others, a recent and particularly interesting application of this type of approach – albeit with an improved treatment of the dynamics of vacancy motion treated at an intermediate scale via the use of fast Fourier transform operators – concerns the predictions of the vacancy-assisted climb in $\gamma\gamma'$ Ni superalloy single crystals. The study shows that in regimes mimicking in-service conditions (e.g., uniaxial stress in the order of 180 MPa and temperatures in the order of 1100 K), dislocation climb can significantly affect the overall dissipation within the system by providing more energetically favorable paths for dislocation to glide. While these studies are up to now limited to relatively short time scale and idealized scenarios, they are likely to be the bedrock for the development of temperature-sensitive constitutive models in which the contribution of climb is not obtained by reverse engineering parameters from experimental data.

## 5 Advances in Sub-scale Physics

## 5.1 Dislocation/Interface Interactions

As stated in the above, DDD has only recently been applied to simulate plasticity in polycrystals (Balint et al. 2008; de Sansal et al. 2010). In these early attempts, grain boundaries have typically been treated as impenetrable barriers to dislocation motion. Clearly, these approaches will necessarily overpredict the generation of "backstresses" due to dislocation pile-ups, thereby not yielding quantitative predictions of well-known effects such as the correlation between the inverse of the square root of the grain size and the yield strength, as per the work of Hall and Petch (Balint et al. 2008), Bauschinger phenomenon, etc. As a dislocation interacts in the short range with a material interface, the number of potential outcomes is relatively limited. Indeed the incoming dislocation can either be stopped near the interface, transmitted through the interface, or absorbed. The outcome depends on the interface character, imposed stress state, temperature, and dislocation character itself (Froseth et al. 2004; Capolungo et al. 2007). For example, atomistic simulations applied to hexagonal close-packed magnesium have clearly revealed that a basal screw dislocation can transmit via cross-slip through a $\{10\bar{1}2\}$ twin interface without seeing any change in its core structure (Serra and Bacon 1995). This is not the case if the $\langle a \rangle$ dislocation interacts with a $\{10\bar{1}1\}$ twin boundary. Similarly edge dislocations will dissociate onto the $\{10\bar{1}2\}$ twin interface, leading to the generation of interface disconnections as well as generating new segments in the twin domain.

The development of predictive approaches describing both short- and long-range interactions between dislocations and materials interfaces is particularly challenging. It entails that (i) the stress field within each point in the medium fully

accounts for the effects of heterogeneous elasticity, manifesting itself through image forces, (ii) one can assess under which stress orientation and magnitude do slip transfer occur, (iii) a model can be developed to predict on which slip system would transfer occur, and (iv) a predictive capability can be developed to fully account for the dislocation core changes (i.e., dissociation, shearing, absorption) occurring within the material interface. Further, as stated in the above (v), a material interface can also act as a dislocation source. (iv) and (v) necessarily induce a change in the local interface character (i.e., local misorientation, excess volume, energy). Given the infinite number of material interfaces that can exist in polycrystals, it is unlikely that an exact physical rendition of all these processes can be derived. "Data mining" approaches might, however, allow for simulations of dislocation/interface interactions to be statistically representative. This has been done to some extent to understand grain boundary mobility (Olmsted et al. 2009) as well as to quantify the energy barrier for dislocation transmission (Sangid et al. 2011). Despite this, several advances have been proposed in recent years to address (i)–(iv). The intent is not to provide an exhaustive list but rather a perspective on such a vast topic.

To address (i), both the superposition method (Weygand et al. 2002) and discrete continuous method (DCM) have been used (Lemarchand et al. 2001; Vattré et al. 2014b). The use of the finite element method necessarily poses the question of the mesh resolution near the material interface (Sobie et al. 2014). Interestingly, numerical methods have been proposed to accurately compute the stress state at interface elements (Liu et al. 2016). The DCM is rigorously derived from a variational formulation – such that driving forces on dislocation are exactly related to the energy gradient – and is expected to be more accurate. Further, recent developments of the fast Fourier transform-based discrete continuous approach to the case of heterogenous elasticity should alleviate the issue of the computational burden associated with images forces (Bertin et al. 2015; Bertin and Capolungo 2018).

Challenges (ii)–(iv) have been addressed in the case of coherent interfaces, particularly twin boundaries in hexagonal close pack metals (Fan et al. 2015b). As detailed in what follows, the development of such models requires considerable input from either topological interface models (Hirth and Pond 1996) and/or atomistic simulations. For the sake of illustration, consider the case of a basal dislocation in hexagonal close-packed magnesium interacting in the short range with a twin boundary on the $\{10\bar{1}2\}$ plane. Seminal works (Serra and Bacon 1995; Serra et al. 1999) have allowed one to establish a catalog of potential reactions between glide dislocations and twin boundaries. One notes that such catalog is still evolving. Consider the case of basal $\langle a \rangle$ dislocations. Three such type of dislocations, with Burgers vector $a_1 = \frac{1}{3}[2\bar{1}\bar{1}0]$, $a_2 = \frac{1}{3}[\bar{1}2\bar{1}0]$, and $a_3 = \frac{1}{3}[\bar{1}\bar{1}20]$, can meet the $(0\bar{1}\bar{1}2)$ twin boundary. As stated previously, the interaction between a screw $a_1$ dislocation and the $(0\bar{1}\bar{1}2)$ twin boundary leads to a complete transmission into the twin domain. As such issues (iii) and (iv) do not require any treatment – the DDD framework can compute the driving force on each dislocation. However, (ii) is not trivially addressed as the dislocation could continue its path on either a basal or a prismatic system in the twin domain. To select the most likely outcome, a local maximum dissipation rule can be introduced. The idea is simply to compute

the product of the driving force and velocities for all nodes in contact with the interface for each potential transfer slip system and select the system that maximizes dissipation. In other cases, for example, $a_2$ intersecting the same twin boundary, one needs to further account for the dissociation of the incoming dislocation into the sum of a twinning partial dislocation and a residual dislocation with Burgers vector $b_r = \frac{[\bar{2}113]}{6}$. The residual dislocation can in turn be transferred into several potential slip systems into the twin domain (i.e., pyramidal type I and II), and the maximum dissipation criterion can be used to select the transferred slip system.

In Fan et al. (2015b), the authors used this approach to simulate the effects of twin volume fraction on both yield strength and hardening of a system subjected to monotonic compression. Figure 4 shows the hardening response as a function of twin volume fraction, $f_t$. The blue lines indicate the disconnections formed as a result of dislocation/twin interactions.

The previous example was concerned with fully coherent interfaces (e.g., twin boundaries) in which challenges (iii) and (iv) can be addressed relatively easily. To increase the level of complexity and in correlation with the recent growth in interest in nanolayered metallic composites (Misra et al. 2005), significant effort has been placed in simulating dislocation/hetero-interface interactions. In particular, Cu/Nb nanolayered composites, consisting of an alternating stack of Cu and Nb layers, have been a focal point of investigation. These heterointerfaces intrinsically contain a network of potentially mobile misfit dislocations rendering the interface relatively weak and capable of shearing the core of an incoming dislocation from either side of the interface (Hoagland et al. 2006). A first series of atomistically guided simulations have been proposed whereby the shearing of the core of the dislocation upon intersecting the hetero-interface is taken account (Sobie et al. 2014). Pragmatically, the numerical approach simply consists of removing segments upon arriving into the interface. It has been shown that these considerations best approximate results from atomistic simulations. Further complexity was latter added



**Fig. 4** Engineering stress strain curves of a twinned domain as a function of twin fraction (Reprinted from Fan et al. 2015b)
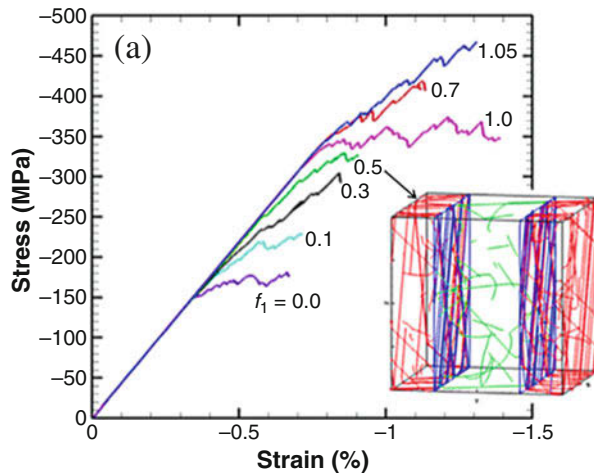
to predict dislocation nucleation from the hetero-interface (Wang et al. 2014). In summary, the authors proposed an original approach whereby both interfacial dislocations and bulk dislocations are modeled simultaneously. Further, interfacial dislocations are allowed to both glide and climb. The climb process, necessarily nonconservative, was further calibrated via atomistic simulations. Nucleation was modeled by essentially letting the interfacial dislocation evolve and simultaneously checking for the alignment of interfacial dislocation lines with potential slip systems, for the driving force acting on said segment and for an energy-based criterion (Frank's rule). These models do not, however, explicitly treat the network of misfit dislocations present at the interface such that substantial calibration against atomistic simulations is required. To circumvent these challenges, recent developments have expanded on the original work of Frank, Bilby, and Bollman in order to predict exactly, and in agreement with atomistic simulations, the network of misfit dislocations within the material interface as well as its dynamic relaxation process and interactions with bulk dislocations (Vattré 2017; Vattré and Pan 2017).

Finally, the case of general boundaries has also received attention. However, to date the solutions proposed do not reflect the complexity of the short-range dislocation/grain boundaries interaction process. To address the barrier for propagation across a boundary, Fan et al. (2015b) simply propose to compute the shear stress at each dislocation in contact with the grain boundary interface and let it transmit once a critical value is reached. The authors proposed to relate the barrier strength to the misorientation angle across the grain boundary $\tau_c = Gbsin^2(\theta_{mis})$, with $\theta_m$ denoting the misorientation angle. Clearly such a criterion does not capture the complexity of the process. In parallel to the aforementioned development and in the restricted case of two-dimensional models, it has been proposed to describe the process of grain boundary sliding by introducing interface dislocations along the grain boundaries as well as to introduce dislocation sources at grain boundaries and triple lines (Quek et al. 2016).

## 5.2    Kinetics

Over the past decade, DDD has been successfully used to, in essence, validate reduced-order models formally relating the change in zero temperature strength of a system to its defect content (e.g., voids, self-interstitial atom loops, precipitates, and other dislocations) (Madec et al. 2003; Greer et al. 2008; Sobie et al. 2015). Recently, however, focus has been placed on unraveling the linkage between both defect content and temperature with the kinetics of slip (e.g., strain rate sensitivity), dislocation nucleation, etc. (Geslin et al. 2017; Sobie et al. 2017a, b, c).

Fundamentally, the local activation of a dislocation-mediated dissipative process (e.g., cross-slip, slip, nucleation) depends both on the system geometry and on the local stress state within the activation volume associated with said process. In an engineering material (i.e., departing from model materials), the microstructure is formally described by the defect content (e.g., the number density of voids and precipitates, dislocation density on each slip system) as well as by the spatial

arrangements between defects. This complexity necessarily raises the question of homogenization, i.e., how to derive laws predicting both the effectiveness and the distributions of the activation of dissipative processes that are statistically representative of bulk behavior?

At the atomistic length scale, both in the context of harmonic transition state theory and of transition state theory in general, a series of algorithms have been proposed to quantify both the minimum energy pathways (MEP) and prefactors (associated with attempt frequencies and entropic contributions) for unit processes (e.g., migration of dumbbells Marinica et al. 2011). These methods include the nudged elastic band (NEB), Dimer, Free End NEB, activation-relaxation technique, umbrella sampling, etc. (Jonsson et al. 1998; Henkelman and Jonsson 2000; Mousseau and Barkema 1998; Saroukhani et al. 2016). Since their inception, MEP search methods have been extensively and successfully used to address point and defect cluster diffusion, dislocation nucleation from material interfaces, dislocation precipitate interactions, etc. (Saroukhani et al. 2016; Zhu et al. 2008). In the context of multiscale modeling and echoing the challenge stated above, the direct application of MEP search methods in the context of atomistic simulations to derive constitutive relationships – or at the very least to calibrate existing laws – will in most cases be inadequate. Indeed, the volumes and geometrical configurations accessible within atomistics do not necessarily satisfy the necessary requirements to render predictions of MEP statistically representative.

Over the past few years, several independent efforts have been made to jointly use DDD and MEP methods (Geslin et al. 2017; Sobie et al. 2017a, b, c). Clearly the significant reduction in degrees of freedom associated with the use of DDD allows one to probe MEPs in statistically representative volumes (in the order of several $\mu m^3$). Two distinct classes of problems have been addressed thus far. First, one can study or quantify the relationship between activation barriers and unit process as a function of geometry and imposed stresses. In this case, one can advantageously perform a direct comparison with atomistic simulations pertaining to smaller systems prior to using DDD to study large volumes. One notes that particularly interesting developments to the DDD framework have been proposed to capture core contributions to the energy landscape in accordance with atomistic simulations (Geslin et al. 2017). Second, the question of coarse graining can be addressed.

While several MEP approaches have been used thus far, here the focus is placed on the NEB method. In short, NEB determines the MEP between transition states following a steepest descent approach. To this end, starting from well-defined initial and final configurations, a set of images are created. These formally represent an initial guess of the trajectory of the system. One such guess can simply consist of a linear interpolation between the initial and final configurations. The NEB algorithm will search for the MEP by solving a modified equation of motion for each image until convergence is reached. As such, all images are simultaneously evolved to equilibrium under a modified force (Eq. (24)) toward a level curve on the energy landscape. The total force on each degree of freedom $i$ is written as (Jonsson et al. 1998)

$$F_i = F_i^s|_\parallel + F_i^R|_\perp, \tag{24}$$

with the true force being related to the gradient of the potential energy within the system,

$$F_i^R = -\nabla V(R_i) = F_i^{PK} . \tag{25}$$

$R$ is the $3N$-dimensional vector containing the coordinates of each dislocation node.

In isotropic elastic media, the force acting on a dislocation is known in closed form from the derivation proposed in Cai and Bulatov (2006). One notes that recent extensions have been proposed in the case of anisotropic media (Po et al. 2018). Finally, the discrete continuous approach – applied either in the context of finite elements or fast Fourier transform methods – provides a direct access to the energy and its gradient in a thermodynamically consistent fashion at any finite element node or voxel, such that a simple interpolation procedure provides access to the force acting on each node without need for the use of an analytically derived solution. Regardless of the numerical approach (i.e., discrete continuous or not), the force acting on a node is projected perpendicularly to the energy gradient:

$$F_i^R|_\perp = F^{PK}(R_i) - F^{PK}(R_i) \cdot \hat{\tau}_i . \tag{26}$$

A modified tangent that depends on local image energy is used (Henkelman and Jonsson 2000) and represents a tangent to the high-dimensional energy landscape. The spring force parallel to the tangent is defined as (Henkelman and Jonsson 2000)

$$F_i^s|_\parallel = k \left[ |R_{i+1} - R_i| - |R_i - R_{i-1}| \right] \hat{\tau}_i . \tag{27}$$

In practice, the NEB method can be coupled to an existing DDD implementation without an onerous numerical implementation. The DDD code naturally calculates the true force in Eq. (24). Each image is evolved similarly to a DDD simulation, except that forces have to be modified according to Eq. (24). The numerical procedure to converge to the MEP is nearly identical to simulating the time evolution of several DDD simulations simultaneously – each image influences its neighbors via the spring force but otherwise evolves independently. Consequently, the method can be readily parallelized with little communication overhead.

For the sake of illustration, consider the unit process associated with the homogeneous nucleation of a dislocation in aluminum. Note here that owing to the low stacking fault energy of the system, novel developments were proposed to fully predict the role of dislocation cores to the nucleation process. Figure 5 shows the initial path assumed with the nucleation of a shear loop (a) as well as the predicted MEP (b) and (c).
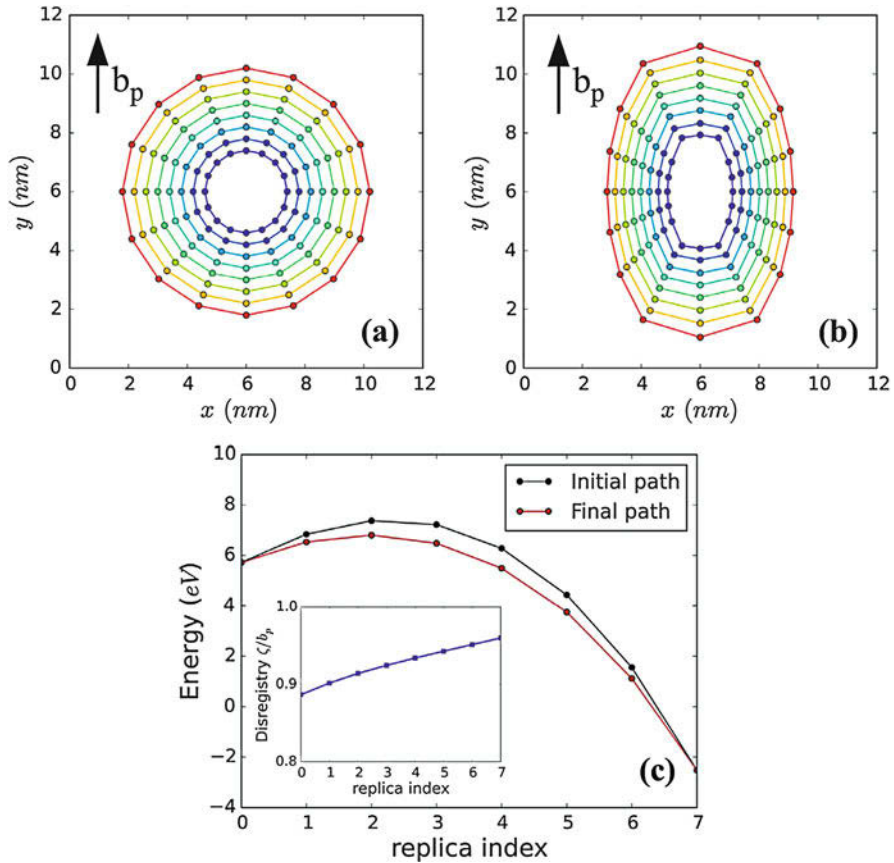
**Fig. 5** Initial configurations used to simulate the homogeneous nucleation of a partial dislocation in Al under an applied resolved shear stress of 1.6 GPA, (**b**) final configuration obtained, (**c**) predictions of the MEP associated with the nucleation process (Reprinted from Geslin et al. 2017)

### 5.2.1 Application to the Case of Dislocation Nucleation

The aforementioned case of homogeneous dislocation nucleation delineates a particularly interesting illustration of the degree to which DDD can be calibrated against atomistic simulations. Pragmatically, the authors (Geslin et al. 2017) proposed to use a discrete continuous method, in the context of the fast Fourier transform method of Bertin et al. (2015), allowing one to have a thermodynamically rigorous relationship between the free energy within the system and the force acting on a dislocation segment or node, depending on the approach chosen. With this, the free energy within the system, $E_{tot}$, can be written as

$$E_{tot} = E_{elastic} + E_{core} + E_{fault}, \tag{28}$$

in which the subscripts "elastic," "core," and "fault" refer to the contributions to the free energy of the system from the linear elastic strain field within the medium, to the core energy of dislocations (in which linear elasticity does not hold), and to the stacking faults within the medium, respectively. All quantities in the above are spatially resolved. With this and in perfect analogy with Eq. (25), the force acting on a dislocation node can be expressed as the negative of the derivative of the total energy within the medium:

$$F_n = -\frac{\partial E_{tot}}{\partial R_n}, \tag{29}$$

in which $F_n$ is the vector force acting on node $n$ and $R_n$ denotes the position vector of node $n$. The contribution to the total force acting on the dislocation due to the elastic strain energy within the medium can be computed either from a direct numerical derivative of $E_{elastic}$ with respect to the position of the node or as a line integral of the Peach-Koehler force over the segments on the left and right of node $n$. One notes that the former method is naturally more computationally involved, in that one typically uses a line integral formally expressed as

$$F_n^{elastic} = \int_{L^n} \frac{s}{L_n} \sigma.b^n \times \xi^n ds + \int_{L^{n+1}} \frac{s}{L_{n+1}} \sigma.b^{n+1} \times \xi^{n+1} ds . \tag{30}$$

Here $\xi^n$ denotes the unit vector tangent to the segment bound by nodes $n-1$ and $n$. Similarly, $L^n$ denotes the length of the segment bound between nodes $n$ and $n+1$.

Following the original proposal from Clouet et al. (2009), the core energy per unit length associated with a dislocation segment can be written as

$$e_{core} = c_0 + c_1\alpha^2 + c_2\alpha^2, \tag{31}$$

in which $\alpha$ introduces the character (i.e., line orientation) dependence of core energy of the dislocation, $\alpha = |b.\xi|/||b||$. With this, the core energy of segment bound by nodes $n-1$ and $n$, $E_n^{core}$, can be written as

$$E_n^{core} = \left(c_0 + c_1\alpha_n^2 + c_2\alpha_n^2\right) L^n, \tag{32}$$

such that the force acting on a dislocation arising from core contributions is simply written as the sum of the partial derivative of $E_n^{core}$ and $E_{n+1}^{core}$ with regard to the position of node $n$. Fortunately, in this case, one can trivially find a closed form expression for such contributions. Finally, in the case of loop nucleation, the contribution of the stacking fault energy due to dislocations can also be taken into account. An approach, similar to that delineated in the above to account for core contributions, is used to this end. The reader is referred to Geslin et al. (2017) for more details.

The approach summarized above shows that the core energy and stacking fault energy contributions can be rigorously calibrated against atomistic simulations, thereby enabling the method to quantitatively predict the activation barriers associated with unit processes. In Geslin et al. (2017), it has been demonstrated that the barrier associated to the nucleation of a shear loop can be predicted with similar accuracy with atomistic of DDD approaches provided that both core and stacking fault energy contributions are taken into account.

### 5.2.2   Connections to Multiscale Modeling

The connection between DDD and MEP search methods allows one to address the key question of homogenization/coarse graining, with particular insight found in seminal papers by McDowell (1997, 1999). Among other things, one can quantify the role of microstructure (i.e., the number density, type, and spatial distribution of defects) on the effective barrier associated with a unit process. Further, from such simulations, one can produce reduced-order models introducing a mathematical linkage between the imposed stress, microstructure, and probability distribution of barriers.

As an example, consider a body center cubic system in which one seeks to quantify the barrier associated with the bypassing of a distribution of $< 100 >$ sessile self-interstitial atom (SIA) loops, as described in Sobie et al. (2017c). Consider first a model that consists of placing an isolated SIA loop within a simulation volume and quantifying the effects of geometry on the predicted activation barriers, which we will refer to as "unit bypass." Periodic boundary conditions are used, creating an array of SIA loops with a mean spacing of 117 nm. (Note that in these simulations, neither the contributions of core energies nor those of the stacking faults are accounted for.) From these simulations, one can extract a distribution of potential barriers, which will be compared to a direct set of simulations in which a dislocation travels through a network of obstacles with similar density and mean spacing as in the unit bypass simulations. As will be detailed in what follows, these approaches, the latter corresponding to the configuration most resembling realistic microstructures, yield vastly different quantification of the barriers encountered by a moving dislocation.

The activation energy for unit bypass in a linear array of SIA loops with spacing of 117 nm was calculated as a function of interaction geometry $(z/L)$ (i.e., the relative position of the glide plane with respect to the loop center, in which $L$ denotes the size of the SIA loop) and an applied shear stress of 10 MPa. The configuration is shown in the left panel of Fig. 6. The activation barriers associated with these processes are shown in the left panel of Fig. 7. One sees a significant effect of the relative position between the glide loop and the SIA loop, showing the inadequacy in quantifying a unique barrier associated with a dislocation-mediated process. Further, the energy shown in the left panel of Fig. 7 can be integrated as a function of relative distance between the glide planes to generate a distribution of barriers akin to what a dislocation would be facing in a more realistic scenario. Assuming an equal likelihood for the relative distance $(z)$ between the glide plane and the SIA loops, the probability of activation barriers can be computed simply by integrating
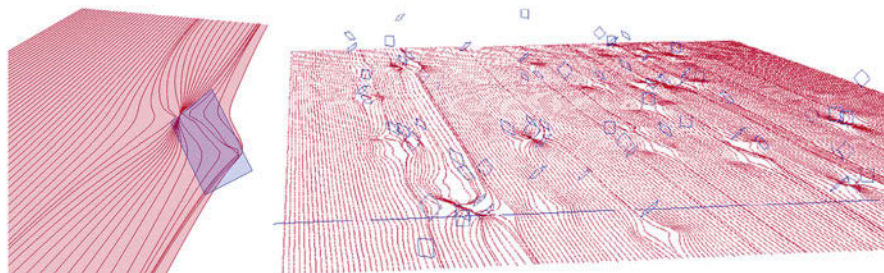
**Fig. 6** (left) The minimum energy pathway taken by a $1/2 \langle 111 \rangle (1\bar{1}0)$ glide dislocation as it overcomes a $\langle 001 \rangle$ SIA loop of 4.3 nm size. Periodic boundary conditions are used such that the mean spacing between the SIA loop array is 117 nm. The red lines correspond to the dislocation configuration as it overcomes the SIA loop. (right) The minimum energy pathway taken by a dislocation as it overcomes an ensemble of SIA loops. The configurations in the left and right panels have similar SIA density and mean spacing. The resolved shear stress imposed in both cases is 10 MPa. (From Sobie et al. 2017b)
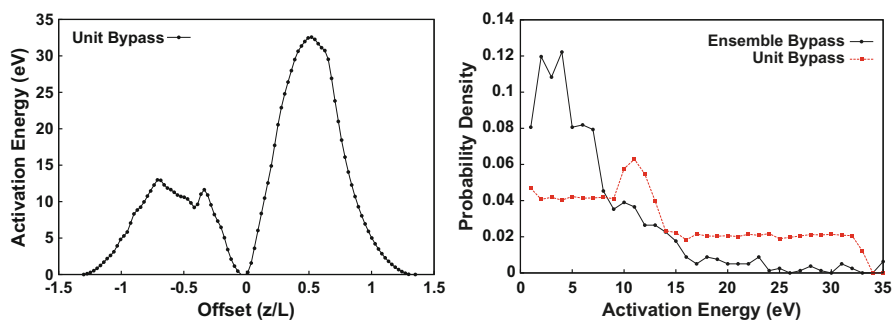


**Fig. 7** (left) Activation barrier corresponding to the bypass of a regular array of SIA loops as a function of the ratio of the distance between the glide plane and center of the SIA loop, $z$, and the SIA loop size is $L$. The energy obtained corresponds to configuration in the left panel of Fig. 6. (right) The black line corresponds to the probability distribution of the activation energy for an edge dislocation bypassing an ensemble of obstacles similar to that depicted in Fig. 6 (right). The red line is the probability distribution obtained by integrating the activation barriers on the left figure (corresponding to unit bypass) over $-1.5$ to $1.5 \, z/L$. (From Sobie et al. 2017c)

the results shown in Fig. 6 (left) for $z/L$ varying from $-1.5$ to $1.5$. The resulting distribution is shown in red in Fig. 7 (right).

The distribution of barriers from the unit bypass simulations can be compared to simulations corresponding to the scenario depicted in Fig. 6 (right). The distribution of barriers in Fig. 6 (right) was obtained by considering a volume containing six randomly positioned SIA loops with a mean separation distance equal to the linear spacing of the array of SIA loops in the previous calculation, 117 nm. The activation energy barriers for glide were calculated for 800 simulation repetitions. The resulting probability density function of activation barriers is shown in Fig. 7 (right panel, black line).

The activation energy probability distribution for an ensemble of loops has fewer defined features than the distribution obtained for unit bypass. The variation in the connected segment lengths and relative defect geometry acts to spread the well-defined peak seen for the unit bypass process. Furthermore, the mean activation energy is 43.4% lower for an ensemble of defects (6.56 vs. 11.6 eV). It is clear that the highly localized interaction between the defects and the dislocation does not entirely characterize the energetics of the process. The change in morphology of the extended dislocation segments neighboring the bypassed defect induces important changes in the activation energy that cannot be captured in unit bypass simulations.

# 6       Summary and Conclusion

In this short summary of recent advances in the field of discrete dislocation dynamics, a few promising developments have been presented. First, from the computational material science viewpoint, two distinct efforts have been proposed to limit the computational burden as well as to increase the effective time step used to solve the dynamics problem. To this end the subcycling method, allowing in essence the reduction of the number of force calculations to be performed within the subcycles, and the fast Fourier transform (FFT) algorithms have been proposed. As is often the case, these two approaches have different advantages. While the FFT method is expected to be far less efficient for smaller dislocation densities (e.g., on the order of $1 \times 10^{12} \, \mathrm{m}^{-2}$), the FFT-based approach, combined with the discrete continuous method, is relatively insensitive to dislocation content and should continue to be effective at the high dislocation densities associated with well-worked or fatigue metals. Moreover, the FFT approach is amenable to the treatment of plasticity in heterogeneous media (i.e., polycrystals, crystals containing precipitates, etc.). However, FFT methods will likely break down if large strains could be reached. Despite these key developments, state-of-the-art methods remain confined to the realm of dynamical loading or, in the case of creep-type simulations, to relatively limited simulation times.

A question that may arise lies in the potential of current modeling capabilities in performing virtual characterization of statistically representative microstructures. To this end, both finite element-based and FFT-based formalisms have been proposed to solve the boundary value problem. One notes that the use of the discrete continuous method advantageously yields thermodynamically consistent stress fields and driving forces acting on dislocations. Clearly, the problem of the treatment of short-range interactions between dislocations and materials interfaces, with varying degree of coherency, is now becoming central to future applications. In this regard, some strategies have been proposed to accurately depict dislocation transmission, core changes, etc. at coherent boundaries. While of great interest and appropriate in the case of coherent twin boundaries, one relies on the development of an atomistic simulation-based dictionary of potential reactions. Clearly, this approach is unlikely to be applicable to general grain boundaries for which a data informatics-driven approach is likely to yield more statistically meaningful simulations. In addition to

treating the effects of grain sizes and texture, a solid interest has been gained in addressing the problem of coupling with the materials chemistry. It is the authors opinion that such a path is likely to be particularly significant as it will allow one to depart from predictions limited to model material systems. While of interest, current approaches are limited to consideration of the vacancy flux toward dislocations, and a complete coupling between both the mechanical and diffusional problems has not yet been proposed.

Finally, in another recent, and promising, advance, DDD is used to extract activation barriers pertaining to unit processes or to sequences of unit processes. To this end, minimum energy pathway search algorithms, originally derived in the context of atomistic simulations, have been adapted to the DDD framework. It has been shown that the DDD formalism can be enhanced to account for the contribution of both dislocation cores and stacking faults enabling the reproduction of energy barriers computed by atomistic simulations. Further, taking the example of dislocation sweeping through a network of obstacles, the inadequacy of extracting activation barriers from non-statistically representative simulations was quantified. These methods open a path toward homogenization of the response of a material from the viewpoint of the energy landscape (or barriers) to be able to derive temperature-sensitive constitutive laws from discrete dislocation dynamics. It must be acknowledged, however, that most DDD simulations rely on the use of overdamped equations of motion, which calls into question the temperature sensitivity of the simulations and, in particular, the accuracy of the prediction of the rate of activation of cross-slip events. One could envisage the use of Langevin dynamics to capture statistical fluctuations in the system that control the activation of rare events.

# References

Anderson PM, Hirth JP, Lothe J (2017) Theory of dislocations, 3rd edn. Cambridge University Press, New York

Arsenlis A, Cai W, Tang M, Rhee M, Oppelstrup T, Hommes G, Pierce TG, Bulatov VV (2007) Enabling strain hardening simulations with dislocation dynamics. Model Simul Mater Sci Eng 15:554–595

Aubry S, Arsenlis A (2013) Use of spherical harmonics for dislocation dynamics in anisotropic elastic media. Model Simul Mater Sci Eng 21:065013

Bacon DJ (1992) Dislocations in crystals. In: Gerold Va (ed) Materials science and technology: a comprehensive treatment, vol 1. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 411–482

Bako B, Clouet E, Dupuy LM, Bletry M (2011) Dislocation dynamics simulations with climb: kinetics of dislocation loop coarsening controlled by bulk diffusion. Philos Mag 91:3173–3191

Balint DS, Deshpande VS, Needleman A, Van der Giessen E (2008) Discrete dislocation plasticity analysis of the grain size dependence of the flow strength of polycrystals. Int J Plast 24:2149–2172

Bertin N, Capolungo L (2018) A FFT-based formulation for discrete dislocation dynamics in heterogeneous media. J Comput Phys 355(Supplement C):366–384

Bertin N, Upadhyay MV, Pradalier C, Capolungo L (2015) A FFT-based formulation for efficient mechanical fields computation in isotropic and anisotropic periodic discrete dislocation dynamics. Model Simul Mater Sci Eng 56:065009

Braisford A, Bullough R (1981) The theory of sink strengths. Philos Trans R Soc Lond Ser A Math Phys Sci 302:87–137

Bulatov VV, Cai W (2006) Computer simulations of dislocations. Oxford, New York

Bulatov VV, Hsiung LL, Tang M, Arsenlis A, Bartelt MC, Cai W, Florando JN, Hiratani M, Rhee M, Hommes G, Pierce TG, Diaz de la Rubia T (2006) Dislocation multi-junctions and strain hardening. Nature 440:1174–1178

Cai W, Bulatov VV (2006) A non-singular continuum theory of dislocations. J Mech Phys Solids 54:561–587

Capolungo L, Spearot D, Cherkaoui M, McDowell D, Qu J, Jacob K (2007) Dislocation nucleation from bicrystal interfaces and grain boundary ledges: relationship to nanocrystalline deformation. J Mech Phys Solids 55:2300–2327

Chaussidon J, Robertson C, Rodney D, Fivel M (2008) Dislocation dynamics simulations of plasticity in fe laths at low temperature. Acta Mater 56:5466–5476

Clouet E, Ventelon L, Willaime F (2009) Dislocation core energies and core fields from first principles. Phys Rev Lett 102:055502

Danas K, Deshpande VS (2013) Plane-strain discrete dislocation plasticity with climb-assisted glide motion of dislocations. Model Simul Mater Sci Eng 21:045008

de Sansal C, Devincre B, Kubin L (2010) Grain size strengthening in microcrystalline copper: a three-dimensional dislocation dynamics simulation. In: Mechanical properties of solids XI. Key engineering materials, vol 423. Trans Tech Publications, Uetikon-Zuerich, pp 25–32

Devincre B, Kubin LP (1997) Mesoscopic simulations of dislocations and plasticity. Mater Sci Eng A 234–236:8–14

de Wit R (1960) The continuum theory of stationary dislocations. Solid State Phys 10:249–292

Fan H, Aubry S, Arsenlis A, El-Awady JA (2015a) Orientation influence on grain size effects in ultrafine-grained magnesium. Scr Mater 97:25–28

Fan H, Aubry S, Arsenlis A, El-Awady JA (2015b) The role of twinning deformation on the hardening response of polycrystalline magnesium from discrete dislocation dynamics simulations. Acta Mater 92:126–139

Fan H, Aubry S, Arsenlis A, El-Awady JA (2016) Grain size effects on dislocation and twinning mediated plasticity in magnesium. Scr Mater 112:50–52

Fivel M (2008a) Discrete dislocation dynamics: principles and recent applications. In: Cazacu O (ed) Multiscale modeling of heterogenous materials: from microstructure to macro-scale properties. Wiley, New York, pp 17–36

Fivel MC (2008b) Discrete dislocation dynamics: an important recent break-through in the modelling of dislocation collective behaviour. Comptes Rendus Physique 9:427–436

Froseth A, Derlet P, Swygenhoven HV (2004) Dislocations emitted from nanocrystalline grain boundaries: nucleation and splitting distance. Acta Mater 52:5863–5870

Frost H, Ashby M (1982) Deformation mechanism maps: the plasticity and creep of metals and ceramics. Pergamon Press, Oxford

Gao S, Fivel M, Ma A, Hartmaier A (2017) 3D discrete dislocation dynamics study of creep behavior in ni-base single crystal superalloys by a combined dislocation climb and vacancy diffusion model. J Mech Phys Solids 102:209–223

Gardner DJ, Woodward CS, Reynolds DR, Hommes G, Aubry S, Arsenlis A (2015) Implicit integration methods for dislocation dynamics. Model Simul Mater Sci Eng 23:025006

Geiser J (2009) Decomposition methods for differential equations. CRC Press, Boca Raton

Geslin PA, Gatti R, Devincre B, Rodney D (2017) Implementation of the nudged elastic band method in a dislocation dynamics formalism: application to dislocation nucleation. J Mech Phys Solids 108:49–67

Ghoniem NM, Sun LZ (1999) Fast-sum method for the elastic field off three-dimensional dislocation ensembles. Phys Rev B 60:128–140

Ghoniem N, Tong S, Sun L (2000) Parametric dislocation dynamics: a thermodynamics-based approach to investigations of mesoscopic plastic deformation. Phys Rev B 61:913–927

Graham JT, Rollett AD, LeSar R (2016) Fast-fourier transform discrete dislocation dynamics. Model Simul Mater Sci Eng 24:085005

Graham JT, LeSar R, Capolungo L (2019, in preparation) Discrete dislocation dynamics based polycrystal plasticity

Greer JR, Weinberger CR, Cai W (2008) Comparing the strength of f.c.c. and b.c.c. sub-micrometer pillars: compression experiments and dislocation dynamics simulations. Mater Sci Eng A 493:21–25

Henkelman G, Jonsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 113:9978–9985

Hirth J, Pond R (1996) Steps, dislocations and disconnections as interface defects relating to structure and phase transformations. Acta Mater 44:4749–4763

Hirth JP, Zbib HM, Lothe J (1998) Forces on high velocity dislocations. Model Simul Mater Sci Eng 6:165–169

Hoagland RG, Hirth JP, Misra A (2006) On the role of weak interfaces in blocking slip in nanoscale layered composites. Philos Mag 86:3537–3558

Hull D, Bacon DJ (2001) Introduction to dislocations, 4th edn. Butterworth Heinemann, Oxford

Jonsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Berne BJ, Ciccotti G, Coker DF (eds.) Classical and Quantum Dynamics in Condensed Phase Simulations, World scientific, Singapore, pp 385–404

Keralavarma S, Benzerga A (2015) High-temperature discrete dislocation plasticity. J Mech Phys Solids 82:1–22

Keralavarma SM, Cagin T, Arsenlis A, Benzerga AA (2012) Power-law creep from discrete dislocation dynamics. Phys Rev Lett 109:265504

Kombaiah B, Murty KL (2015) High temperature creep and deformation microstructures in recrystallized zircaloy-4. Philos Mag B 95:1656–1679

Kubin LP (2013) Dislocations, mesoscale simulations and plastic flow. Oxford University Press, Oxford

Kubin LP, Canova G (1992) The modelling of dislocation patterns. Scr Met Mater 27:957–962

Lebensohn RA (2001) N-site modeling of a 3D viscoplastic polycrystal using fast fourier transform. Acta Mater 49:2723–2737

Lebensohn RA, Kanjarla KA, Eisenlohr P (2012) An elasto-viscoplastic formulation based on fast fourier transforms for the prediction of micromechanical fields in polycrystalline materials. Int J Plast 32–33:59–69

Lemarchand C, Devincre B, Kubin LP (2001) Homogenization method for a discrete-continuum simulation of dislocation dynamics. J Mech Phys Solids 49:1969–1982

LeSar R (2014) Simulations of dislocation structure and response. Ann Rev Condens Matter Phys 5:375–407. https://doi.org/10.1146/annurev-conmatphys-031113-133858

Liu B, Raabe D, Roters F, Eisenlohr P, Lebensohn RA (2010) Comparison of finite element and fast Fourier transform crystal plasticity solvers for texture prediction. Model Simul Mater Sci Eng 18:085005

Liu B, Arsenlis A, Aubry S (2016) Computing forces on interface elements exerted by dislocations in an elastically anisotropic crystalline material. Model Simul Mater Sci Eng 24:055013

Madec R, Devincre B, Kubin L, Hoc T, Rodney D (2003) The role of collinear interaction in dislocation-induced hardening. Science 301(5641):1879–1882

Marinica MC, Willaime F, Mousseau N (2011) Energy landscape of small clusters of self-interstitial dumbbells in iron. Phys Rev B 83:094119

McDowell DL (1997) Evolving structure and internal state variables. Nadai award lecture. ASME IMECE, Dallas

McDowell DL (1999) Non-associative aspects of multiscale evolutionary phenomena. In: Picu R, Krempl E (eds) Proceedings 4th international conference on constitutive laws for engineering materials. Rensselaer Polytechnic Institute, Troy, pp 54–57

Michel JC, Moulinec H, Suquet P (1999) Effective properties of composite materials with periodic microstructure: a computational approach. Comput Methods Appl Mech Eng 172:109–143

Misra A, Hirth J, Hoagland R (2005) Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites. Acta Mater 53:4817–4824

Mordehai D, Clouet E, Fivel M, Verdier M (2008) Introducing dislocation climb by bulk diffusion in discrete dislocation dynamics. Philos Mag 88:899–925

Morrow BM, Anderson KR, Kozar RW, Mills M (2013) An examination of the use of the modified jogged-screw model for predicting creep behavior in zircaloy-4. Acta Inf 61:4452–4460

Moulinec H, Suquet P (1998) A numerical method for computing the overall response of nonlinear composites with complex microstructure. Comput Methods Appl Mech Eng 157:69–94

Mousseau N, Barkema GT (1998) Traveling through potential energy landscapes of disordered materials: the activation-relaxation technique. Phys Rev E 57:2419–2424

Mura T (1987) Micromechanics of defects in solids. Martinus Nijhoff, Boston

Olmsted DL, Holm EA, Foiles SM (2009) Survey of computed grain boundary properties in face-centered cubic metals090000ii: grain boundary mobility. Acta Mater 57:3704–3713

Po G, Lazar M, Chandra Admal N, Ghoniem N (2018) A non-singular theory of dislocations in anisotropic crystals. Int J Pasticity 103:1–22

Prasad Reddy GV, Robertson C, Depres C, Fivel M (2013) Effect of grain disorientation on early fatigue crack propagation in face-centred-cubic polycrystals: a three-dimensional dislocation dynamics investigation. Acta Materialia 61:5300–5310

Quek SS, Wu ZX, Zhang YW, Srolovitz DJ (2014) Polycrystal deformation in a discrete dislocation dynamics framework. Acta Mater 75:92–105

Quek SS, Chooi ZH, Wu Z, Zhang YW, Srolovitz DJ (2016) The inverse hall-petch relation in nanocrystalline metals: a discrete dislocation dynamics analysis. J Mech Phys Solids 88(Supplement C):252–266

Sangid MD, Ezaz T, Sehitoglu H, Robertson IM (2011) Energy of slip transmission and nucleation at grain boundaries. Acta Mater 59:283–296

Saroukhani S, Nguyen LD, Leung KWK, Singh CV, Warner DH (2016) Harnessing atomistic simulations to predict the rate at which dislocations overcome obstacles. J Mech Phys Solids 90:203–214

Serra A, Bacon D (1995) Computer simulation of screw dislocation interactions with twin boundaries in h.c.p. metals. Acta Met Mater 43:4465–4481

Serra A, Bacon D, Pond R (1999) Dislocations in interfaces in the h.c.p. metals090000i. Defects formed by absorption of crystal dislocations. Acta Mater 47:1425–1439

Sills RB, Cai W (2014) Efficient time integration in dislocation dynamics. Model Simul Mater Sci Eng 22:025003

Sills RB, Aghaei A, Cai W (2016a) Advanced time integration algorithms for dislocation dynamics simulations of work hardening. Model Simul Mater Sci Eng 24:045019

Sills RB, Kuykendall WP, A AA, Cai W (2016b) Fundamentals of dislocation dynamics simulations. In: Weinberger CR, Tucker GJ (eds) Multiscale materials modeling for nanomechanics. Springer, Cham, p 5317

Siška F, Weygand D, Forest S, Gumbsch P (2009) Comparison of mechanical behaviour of thin film simulated by discrete dislocation dynamics and continuum crystal plasticity. Comput Mater Sci 45:793–799

Sobie C, McPhie MG, Capolungo L, Cherkaoui M (2014) The effect of interfaces on the mechanical behaviour of multilayered metallic laminates. Model Simul Mater Sci Eng 22:045007

Sobie C, Bertin N, Capolungo L (2015) Analysis of obstacle hardening models using dislocation dynamics: application to irradiation-induced defects. Met Mater Trans A 46:3761–3772

Sobie C, Capolungo L, McDowell DL, Martinez E (2017a) Modal analysis of dislocation vibration and reaction attempt frequency. Acta Mater 134:203–210

Sobie C, Capolungo L, McDowell DL, Martinez E (2017b) Scale transition using dislocation dynamics and the nudged elastic band method. J Mech Phys Solids 105:161–178

Sobie C, Capolungo L, McDowell DL, Martinez E (2017c) Thermal activation of dislocations in large scale obstacle bypass. J Mech Phys Solids 105:150–160

Vattré A (2017) Elastic strain relaxation in interfacial dislocation patterns: a parametric energy-based framework. J Mech Phys Solids 105(Supplement C):254–282

Vattré A, Pan EN (2017) Interaction between semicoherent interfaces and volterra-type dislocations in dissimilar anisotropic materials. J Mater Res 32:3947–3957

Vattré A, Devincre B, Feyel F, Gatti R, Groh S, Jamond O, Roos A (2014a) Modelling crystal plasticity by 3d dislocation dynamics and the finite element method: the discrete-continuous model revisited. J Mech Phys Solids 63:491–505

Vattré AJ, Abdolrahim N, Kolluri K, Demkowicz MJ (2014b) Computational design of patterned interfaces using reduced order models. Nat Sci Rep 4:1

Verdier M, Fivel M, Groma I (1998) Mesoscopic scale simulation of dislocation dynamics in fcc metals: principles and applications. Model Simul Mater Sci Eng 6:755–770

Wang HY, LeSar R (1995) O(N) algorithm for dislocation dynamics. Philos Mag A 71:149–164

Wang J, Zhou C, Beyerlein IJ, Shao S (2014) Modeling interface-dominated mechanical behavior of nanolayered crystalline composites. JOM 66:102–113

Wang Z, Ghoniem NM, Swaminarayan S, LeSar R (2006) A parallel algorithm for 3D dislocation dynamics. J Comput Phys 219:608–621

Wang ZQ, Beyerlein IJ, LeSar R (2007) Dislocation motion in high-strain-rate deformation. Philos Mag 87(16):2263–2279

Weygand D, Friedman LH, der Giessen EV, Needleman A (2002) Aspects of boundary-value problem solutions with three-dimensional dislocation dynamics. Model Simul Mater Sci Eng 10:437

Yin J, Barnett DM, Cai W (2010) Efficient computation of forces on dislocation segments in anisotropic elasticity. Model Simul Mater Sci Eng 18:045013

Zbib HM, Diaz de la Rubia T, Rhee M, Hirth JP (2000) 3D dislocation dynamics: stress-strain behavior and hardening mechanisms in fcc and bcc metals. J Nucl Mater 276:154–165

Zheng Z, Balint DS, Dunne FPE (2016) Discrete dislocation and crystal plasticity analyses of load shedding in polycrystalline titanium alloys. Int J Plast 87:15–31

Zhou CZ, LeSar R (2012) Dislocation dynamics simulations of plasticity in polycrystalline thin films. Int J Plast 30–31:185–201

Zhu T, Li J, Samanta A, Leach A, Gall K (2008) Temperature and strain-rate dependence of surface dislocation nucleation. Phys Rev Lett 100:025502

# Mesoscale, Microstructure-Sensitive Modeling for Interface-Dominated, Nanostructured Materials

Irene J. Beyerlein and Marko Knezevic

## Contents

## Abstract

Predictions of the mechanical response of nanocrystalline metals and underlying microstructure evolution and deformation mechanisms are critically important for the manufacturing and design of new advanced structural metals that aim to

I. J. Beyerlein (✉)
Mechanical Engineering Department, Materials Department, University of California, Santa Barbara, CA, USA
e-mail: beyerlein@ucsb.edu

M. Knezevic
Department of Mechanical Engineering, University of New Hampshire, Durham, NH, USA

outperform those in use today. In this chapter, recent advancements in modeling processing-microstructure-property relationships of nanocrystalline metals are covered. These developments include linking mesoscopic microstructure, such as grain (orientation, size, and shape), grain boundaries, twin boundaries, and interface properties with the development of local stress states and deformation mechanisms during mechanical processing or straining. Many recent examples of these techniques are discussed, particularly those demonstrating unanticipated couplings between size effects and texture development. The chapter concludes with a discussion of recommended directions and associated challenges to further advance mesoscale modeling efforts.

## 1 Introduction

Many future engineering systems will rely on high-performance metals that bear strengths and toughness values several times of those in use today. In several situations, these superior properties will be desired in harsh environments, such as elevated temperatures, at high rates, and under radiation. Nanolaminates, built from stacks of crystalline layers each with nanoscale individual thickness, are proving to exhibit many of these target properties. Examples span from nanotwinned materials, wherein individual nanolayers alternate in orientation according to the twin-matrix relationship and are divided by a twinned interface, to biphase nanolaminates, in which the nanolayers differ not only in orientation but also in chemistry and crystal structure (Dalla Torre et al. 2002; Ebrahimi et al. 1999; Hughes et al. 1986; Schuh et al. 2002; Schwaiger et al. 2003; El-Sherik et al. 1992; Xiao et al. 2001; Godon et al. 2010; Champion et al. 2001; Chen et al. 2006; Das et al. 2006; Khan et al. 2008; Shen et al. 1995; Youngdahl et al. 1997; Youssef et al. 2004; Nizolek et al. 2016). Nanotwinned materials have been shown to exhibit excellent strength and hardness, even harder than their nanocrystalline counterparts, while exhibiting reasonable ductility (Lu et al. 2004, 2009a; Beyerlein et al. 2014), outstanding thermal stability (Zhang et al. 2005; Anderoglu et al. 2008), electrical conductivity (Chen et al. 2007; Zhang et al. 2007), and fatigue resistance (Pan et al. 2013; Pan and Lu 2014). The number of superior structural properties biphase nanolaminates is equally large, including strengths that are over five to ten times higher and hardness values that are several orders of magnitude larger than values expected from volume averaged properties of their constituents (Beyerlein et al. 2013a; Mara et al. 2008; Mara and Beyerlein 2014; Han et al. 2013a; Misra et al. 2004, 2007). While the properties are clearly attractive, use of these materials in application relies on understanding and predicting these new properties based on

known and measureable variables, such as basic nanostructure dimensions, chemical composition, and loading conditions, including temperature and strain rate.

All structural metals are polycrystalline, or aggregates of grains, wherein each grain has a distinct crystallographic orientation with respect to all its nearest neighboring grains (Barrett and Massalski 1966). In conventional polycrystalline materials, the sizes of the individual grains have micron and submicron dimension. In the nanostructured materials, these grains have nanoscale dimension, and hence the density of internal boundaries, such as grain, twin, and biphase boundaries, is unconventionally large.

Many ways are being employed to make these nanostructured materials, from bottom-up deposition to top-down forming processes (Yasuna et al. 2000; Carpenter et al. 2015a, b; Kikuchi et al. 1997; Nizolek et al. 2015; Beyerlein et al. 2013b, c; Wang et al. 2013; Kang et al. 2012). As commonly expected of metals, their final nanostructure is sensitive to the how the material was made. Quantifying both the processing/nanostructure relationships and nanostructure-property relationships constitutes a multiscale modeling challenge. Multiscale models for conventional coarse-grained polycrystalline materials have been in development for several years, but analogous versions for nanostructured aggregates require extensions to explicitly account for the effect of internal boundaries on these relationships.

Many multiscale mechanics models for understanding and calculating the deformation of polycrystalline employ a combination of crystal plasticity (CP) theory and polycrystal modeling schemes for which there are many. CP theory relates the distortion of a strained crystal to slip on crystallographic slip systems (Kocks et al. 1998). Polycrystal plasticity models link individual grain response predicted by CP theory to the overall mechanical response of a polycrystalline aggregate.

These polycrystal plasticity models appear in various levels of sophistication and computational efficiency. Among the more popular are mean-field approaches, such as the full constraints Taylor model (1938), and self-consistent schemes such as viscoplastic self-consistent (VPSC) and elastoplastic self-consistent (EPSC) (Lebensohn and Tomé 1993, 1994; Lebensohn et al. 2007), which homogenize the grain neighborhood of an individual grain. Used less often are crystal plasticity models that spatially resolve the grain neighborhoods, which can be referred to as 3D full-field, spatially resolved mechanics techniques. In this class, there are two main types of mesoscale models differentiated by their solution technique: CP finite element (FE) models (Bronkhorst et al. 1991, 1992) and Green's function fast Fourier transform (FFT) models (Lebensohn et al. 2004, 2011; Lieberman et al. 2016; Liu et al. 2010).

Compared to mean-field approaches, 3D full-field, spatially resolved can yield additional information on the effects of grain-grain interactions, intragranular stress and strain evolution, and heterogeneous onset of localization, all of which are important for understanding and designing nanocrystalline materials, as will be demonstrated in this chapter. They also account for local neighborhoods and spatial resolution in the mechanical fields below the grain scale, such as stress concentrations at boundaries and intersections of many boundaries (e.g., triple points, twin/grain boundary intersections). These capabilities are important for modeling

the many boundary-driven mechanisms that are sensitive to a combination of inter-face morphology, character, and crystal orientation, as will be shown later as well.

This chapter will cover studies that have employed and extended mesoscale 3D, full-field spatial resolved computational models to investigate processing-microstructure relationships in boundary-dominant materials. These include examining the influence of bimetal interfaces on slip activity in adjoining grains and on texture evolution. Also presented and discussed here are advancements made to this type of models to include the effects of grain size and shape and layer and twin thickness on dislocation emission and propagation.

## 2    Part I: Methodology

### 2.1    Multiscale Overview

As mentioned, a 3D spatially resolved scheme used frequently in mechanical deformation modeling is the crystal plasticity finite element (CPFE) method. This technique can provide predictions of not only texture evolution but also evolution of intragrain and intergrain misorientations, grain shape, and grain boundary character distribution.

An abbreviated review of the formulation of CPFE is provided below. Regarding notation, vector and tensors are indicated by boldfaced characters and are not italicized. For a second-order tensor $\mathbf{A}$, the notation $\mathbf{A}^T$ denotes its transpose, i.e., $A_{ij}{}^T = A_{ji}$ for $i, j = 1 \dots 3$. The notation tr$\mathbf{A}$ means its trace, while det$\mathbf{A}$ signifies its determinant. The contracted product between two symmetric tensors, $\mathbf{A}$ and $\mathbf{B}$, is defined as: $\mathbf{A} \cdot \mathbf{B}$. The dyadic product of any two vectors, $\mathbf{a}$ and $\mathbf{b}$, is a second-order tensor given by $\mathbf{a} \otimes \mathbf{b} = a_i b_j$ Scalars and tensor components are italicized and not boldfaced.

Figure 1 displays a multiscale CPFE model and the submodel components that it links. Going from the left to right side (frames a–c), the material length increases, and each frame represents a specific instance at which the material response is being evaluated. At the coarsest level (frame a) is the material response of a polycrystal, which is in the form of a granular microstructural model, by the use of the finite element (FE) homogenization method. As a full-field model, this method fulfills both stress equilibrium and strain compatibility conditions making it a suitable modeling tool for capturing the interactions between the constituent grains. Each grain in the polycrystal is represented by an element set, which discretizes a given grain into finite elements (frame b). At this length scale, at each FE integration point, the material constitutive response is being estimated utilizing crystal plasticity (CP) theory. In application of CP, the model usually allows the strain to be accommodated by the simultaneous action of crystallographic slip and deformation twinning. In CPFE, the shear accommodated by the latter mechanism is most often modeled as slip and is referred to as the "pseudo slip" model for twinning (Van Houtte 1978; Kalidindi 1998).
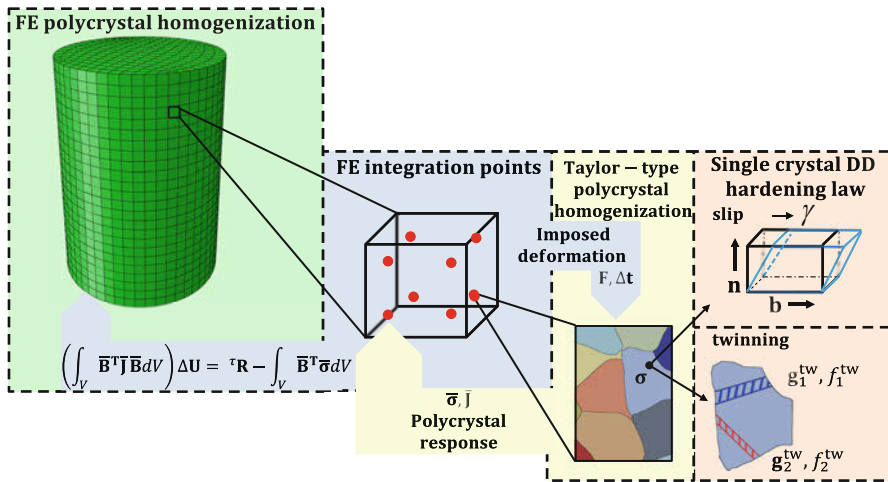
**Fig. 1** A multiscale CPFE modeling framework for the plastic deformation of polycrystals (Taken with permission from Ardeljan et al. 2015a)

## 2.2 Meshing Microstructure, Grain Boundaries, and Interfaces

It is widely recognized that modeling both grain and grain boundary evolution during deformation is critical for understanding material response. Capturing important neighboring grain interactions on deformation is one of the primary reasons for choosing a full-field, spatially resolved over the less computationally expensive mean-field approaches. However, common methods for generating 3D microstructures have been limited in how well they can represent grain morphologies and grain boundaries. In most studies, the grains were blocks or polygons, such as cuboids, rhombic dodecahedrons, and truncated octahedrons. Such inaccurate representations of grain structure and grain boundaries can create numerical artifacts in mechanics calculations of stress and strain states near microstructural boundaries and interfaces (Ritz and Dawson 2008).

Since then, many techniques have been used to better represent microstructures than cuboid grains. One widely used technique for generating 3D microstructures for these codes is the Voronoi tessellation method. The technique does well in creating a set of near-equiaxed grains. The Voronoi tessellation starts with random grain seeds and creates polyhedral-shaped grains. This method, however, still has its limitations. Polyhedral-shaped nonuniform grain shapes created by the Voronoi tessellation methods are often unrealistic because rules for the organization and geometrical constraints of the grains produced are not unique. In addition, the grain boundaries appear as coarse disordered polygons. To produce say elongated grains, it is possible to place these seeds far apart from each other along one dimension in order to get elongated grains, but generally it is difficult to control

the grain aspect ratios. Furthermore, since the grain boundary between the two neighboring domains (grains) is created halfway between the seed points, it proves to be challenging to acquire neighboring grains of dissimilar sizes. Grains often do not look realistic and grain boundaries appear as very coarse polygons, and as a result, the grain boundaries are not continuous surfaces. Finally, a characteristic of Voronoi tessellation is perfectly planar grain boundaries, which generally are notably different from real boundaries inferred from microstructure characterization experiments.

In recent years, explicit meshing of the grain structure and grain boundary surfaces and capturing their evolution with plastic strain have notably advanced the predictive capabilities of these 3D full-field approaches. Creation of realistic grain structures has been largely addressed with the introduction of techniques for explicit grain structure. Gaining much use is DREAM.3D, a software that generates the 3D synthetic voxelized microstructure and the surface meshes for the grain boundaries (DREAM.3D Version 4.2 2013; Groeber and Jackson 2014). It overcomes many of the limitations with the commonly used Voronoi tessellation method. The digital microstructure generated in DREAM.3D appears far more realistic in terms of grain morphology and grain size distribution than that generated by Voronoi tessellation scheme in 3D.

The output from DREAM.3D can be used directly in the voxel-based CP-FFT codes, but since meshing of grain structures is not a capability of the DREAM.3D software, additional processing is needed so that the output can be used to create a 3D mesh for grains and grain boundaries in CPFE. Recent work by Knezevic et al. (2014) and Ardeljan et al. (2015a) developed an integrated tool set that takes a grain boundary surface mesh from DREAM.3D and provides a volume mesh for each 3D grain and its grain boundary surfaces. Before acquiring the triangular surface mesh for each grain, it is convenient to apply a Laplacian-based smoothing filter available in DREAM.3D. The filter smooths out and suppresses any jagged boundaries between grains to achieve better grain structure representation. The final polycrystalline aggregate is comprised of element sets representing individual grains (Fig. 2a, c) and grains with conformal grain boundaries between neighboring 3D grains (Fig. 2b, d). The conformal conditions between constituent grains mean that neighboring grains share triangular elements at grain boundaries.

## 2.3   Modeling Single Crystal Constitutive Response

In order to determine the single crystal response at each integration point, a User MATerial (UMAT) subroutine based on crystal plasticity constitutive formulation is used in Abaqus Standard. The framework facilitates various loading conditions from low to high level of complexity that can be applied in the form of suitable boundary conditions. This applied load is divided into time/strain increments, where for each one, a global stress equilibrium solution is found using a numerically
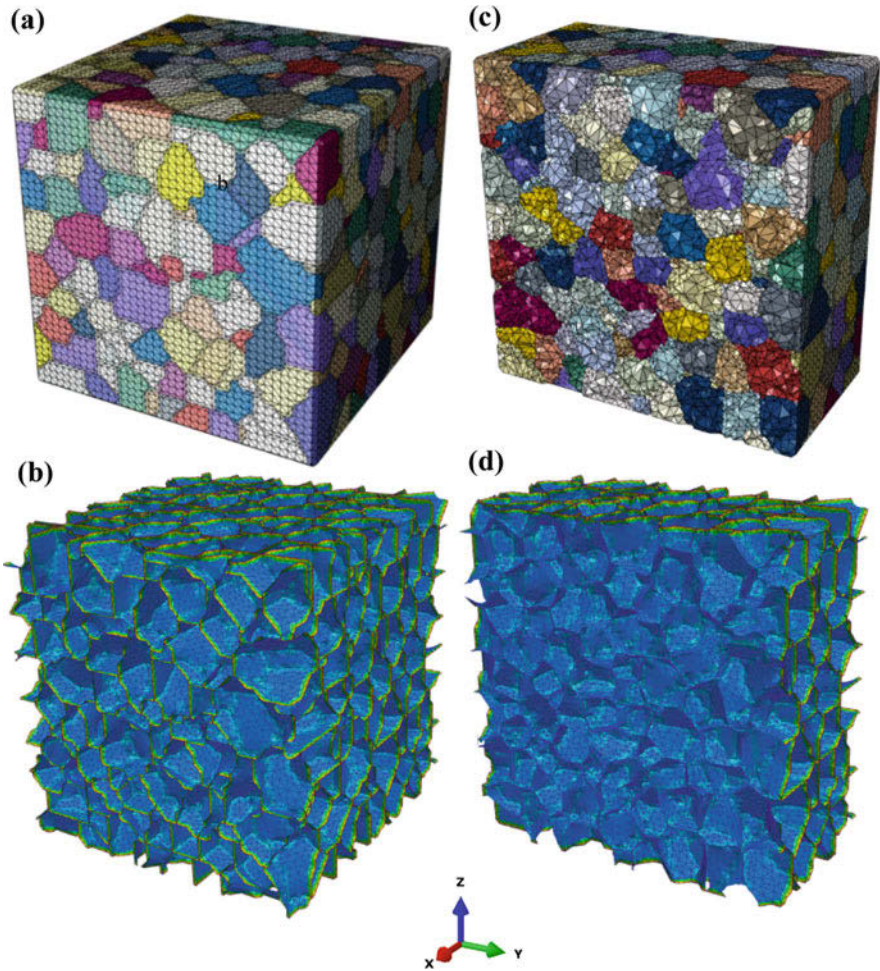
**Fig. 2** An explicit grain structure model consisting of 784 equiaxed grains generated synthetically using DREAM.3D: (**a**) voxel-based model with highlighted edges, (**b**) triangular surface mesh. The triangular surface mesh in (**b**) describes grain boundaries. (**c, d**) The midsection along half the x-axis shows the internal structure of the (**c**) finite element-based model and (**d**) triangular surface mesh. The edge length of the cube is 1000 μm, the average grain size is 135 μm, and the total number of tetrahedral elements (type C3D4 or C3D10 in ABAQUS) in (**c**) is approximately 1,000,000 (Taken with permission from Knezevic et al. 2014)

iterative procedure of the finite element method. This end is achieved by solving the nonlinear FE governing equation in its linearized form, which is given by:

$$\left( \int_V \mathbf{B}^T \mathbf{J} \mathbf{B} dV \right) \Delta \mathbf{U} = \mathbf{R} - \int_V \mathbf{B}^T \boldsymbol{\sigma} dV. \tag{1}$$

In this relation, the listed quantities are, respectively, $\mathbf{B}$, the quantities are, respectively,finite element strain-displacement matrix; $\mathbf{J}$, the material Jacobian matrix; $\mathbf{\Delta U}$, the displacement increment solution; $\mathbf{R}$, the applied force vector; and $\boldsymbol{\sigma}$, the Cauchy stress tensor.

An essential part of CPFE is the crystal plasticity constitutive law that relates the material stress to material distortion (stretch plus rotation) at each integration point within each finite element in the model.

The total velocity gradient tensor, $\mathbf{L}$, can be additively decomposed as the following:

$$\mathbf{L} = \mathbf{L}^e + \mathbf{L}^p, \tag{2}$$

where $\mathbf{L}^e$ and $\mathbf{L}^p$ represent elastic and plastic velocity gradients, respectively. The plastic part of the velocity gradient contains the contributions from both slip and twinning via

$$\mathbf{L}^p = \mathbf{L}^{sl} + \mathbf{L}^{tw}. \tag{3}$$

The corresponding contributions to the velocity gradients, due to slip and twinning, are further expressed as:

$$\begin{aligned} \mathbf{L}^{sl} &= \sum_{\alpha}^{N^{sl}} \dot{\gamma}^{\alpha} \mathbf{m}_o^{\alpha} = \sum_{\alpha}^{N^{sl}} \dot{\gamma}^{\alpha} \mathbf{b}_o^{\alpha} \otimes \mathbf{n}_o^{\alpha}, \\ \mathbf{L}^{tw} &= \sum_{\beta}^{N^{tw}} \dot{f}^{\beta} S^{\beta} \mathbf{m}_o^{\beta} = \sum_{\beta}^{N^{tw}} \dot{f}^{\beta} S^{\beta} \mathbf{b}_o^{\beta} \otimes \mathbf{n}_o^{\beta}, \end{aligned} \tag{4}$$

where $\dot{\gamma}^{\alpha}$ stands for the shearing rate on slip system $\alpha$; $\mathbf{m}_o^{\alpha}$ and $\mathbf{m}_o^{\beta}$ are the Schmid tensors associated with slip system $\alpha$ and twin system $\beta$, respectively; and $S^{\beta}$ is the characteristic twin shear for the twin system $\beta$. Finally, $N^{sl}$ and $N^{tw}$ represent the total number of available slip and twinning systems, respectively. The Schmid tensors represent the unit slip or twin system tensor, defined as the dyadic product ($\otimes$) between the unit Burgers direction ($\mathbf{b}_o$) and unit plane normal ($\mathbf{n}_o$) vectors of slip system $\alpha$ or twin system $\beta$, respectively, in the undeformed configuration indicated by subscript "$o$." The rate of change of the twin volume fraction per twin system $(\dot{f}^{\beta})$ is related to the shear rate on the twin system $(\dot{\gamma}^{\beta})$ according to

$$\dot{f}^{\beta} = \frac{\dot{\gamma}^{\beta}}{S^{\beta}}. \tag{5}$$

In the finite deformation formulation, an assumption is made that the deformation gradient ($\mathbf{F}$) can be multiplicatively decomposed into its elastic ($\mathbf{F}^e$) and plastic ($\mathbf{F}^p$) contributions according to

$$\mathbf{F} = \mathbf{F}^e \mathbf{F}^p, \tag{6}$$

where the elastic component contains contributions from both elastic stretching and lattice rotation, while the plastic component embodies contributions due to

plastic deformation. The constitutive relationship between $\mathbf{F}^e$ and stress in the crystal is obtained considering two expressions for the second Piola-Kirchhoff stress tensor $\mathbf{T}^e$,

$$\mathbf{T}^e = \mathbf{C}\mathbf{E}^e, \, \mathbf{T}^e = \mathbf{F}^{e^{-1}} \left\{ \left( \det \mathbf{F}^e \right) \boldsymbol{\sigma} \right\} \mathbf{F}^{e^{-T}}, \tag{7}$$

where $\mathbf{C}$ is the fourth-order elasticity tensor and $\boldsymbol{\sigma}$ is the Cauchy stress in the crystal. $\mathbf{E}^e$, the Lagrangian finite strain tensor is related to $\mathbf{F}^e$ via

$$\mathbf{E}^e = \frac{1}{2} \left\{ \mathbf{F}^{e^T} \mathbf{F}^e - \mathbf{I} \right\}, \tag{8}$$

Finally to compute stress, the evolution of $\mathbf{F}^p$ needs to be evaluated. It is determined by crystallographic slip and twinning (microshear rates), and it can be expressed in rate form using the following flow-rule relationship:

$$\dot{\mathbf{F}}^p = \mathbf{L}^p \mathbf{F}^p. \tag{9}$$

Integrating Eq. (9) from $t$ to $\tau = t + \Delta t$ yields

$$\mathbf{F}^p(\tau) = \exp\left(\mathbf{L}^p \Delta t\right) \mathbf{F}^p(t). \tag{10}$$

Furthermore, the exponential can be conveniently approximated and further expanded using Eq. (3) to give

$$\mathbf{F}^p(\tau) = \left\{ \mathbf{I} + \Delta t \mathbf{L}^p \right\} \mathbf{F}^p(t) = \left\{ \mathbf{I} + \Delta t \left( \mathbf{L}^{sl} + \mathbf{L}^{tw} \right) \right\} \mathbf{F}^p(t), \tag{11}$$

where again, $\mathbf{I}$ is the identity matrix.

Moreover, the previous equation can be rewritten as

$$\mathbf{F}^{p^{-1}}(\tau) = \mathbf{F}^{p^{-1}}(t) \left\{ \mathbf{I} - \Delta t \left( \mathbf{L}^{sl} + \mathbf{L}^{tw} \right) \right\}. \tag{12}$$

## 2.4    Subgrain Mechanisms: Slip Resistances

Predicting slip activity is fundamental to predicting the evolution of microstructure, textures (the reorientations of the crystals), and material flow stress with strain. Through Eq. (4), the theory of crystal plasticity can relate slip activity on crystallographic slip and twinning systems to the distortion of a crystal. It does not, however, include criteria for the threshold of slip. Nearly all crystal plasticity-based constitutive models, including the three methods presented above, polycrystal models, CPFE, and CP-FFT, require choosing a criterion for activating a slip system in order to predict slip activity.

According to the thermodynamics theory of slip, in the regime of strain rates, roughly from $10^{-5}$/s to $10^{5}$/s, dislocation motion is thermally activated, and dislocations must overcome energetic barriers in order to move. For this regime, the commonly used flow rule of viscoplasticity relating slip rate to resolved stress has a power-law form:

$$\dot{\gamma}^s = \dot{\gamma}_0 \left( \frac{|\tau^s|}{\tau_c^s} \right)^{\frac{1}{m}} \text{sign}\left( \tau^s \right), \ \dot{\gamma}^s = \begin{cases} \dot{\gamma}_0 \left( \frac{|\tau^s|}{\tau_c^s} \right)^{\frac{1}{m}} \text{sign}\left( \tau^s \right) \text{ if } \tau^s > 0 \\ \qquad\qquad 0 \qquad\qquad \text{ if } \tau^s < 0 \end{cases} \tag{13}$$

where $\dot{\gamma}^s$ is the shear strain rate for a slip system $s$. The resolved shear stress is given by $\tau^s = \boldsymbol{\sigma} \cdot \mathbf{m}_0^s$, which is the tensor product of the stress $\sigma$ and Schmid tensor $\mathbf{m}_0^s$, and where $\dot{\gamma}_0$ is a reference slip rate (arbitrarily taken here as 0.001 s$^{-1}$). In the exponent, $m$ denotes the strain rate sensitivity factor. This power-law form is desirable because it provides uniqueness of solution for the active slip systems that accommodate an imposed strain rate.

The flow rule introduces a threshold critical resolved shear stress $\tau_c^s$ in order to activate slip. It is commonly called the critical resolved shear stress (CRSS). Nearly all CRSS models introduced over the past several decades are phenomenological. The simplest model for the CRSS is a constant value, which does not evolve with strain. However, for calculations of deformation behavior beyond the elastic-plastic yield transition, it is desirable that the CRSS represent the resistance to move dislocations, consistent with changes in subcrystalline and granular microstructure. Most CRSS hardening models have been developed for coarse-grained polycrystals, where in the large grains, dislocations tend to accumulate in amounts that increase with strain and depend on strain rate and temperature (Beyerlein and Tomé 2008). CP models that employ these hardening laws are able to predict changes in constitutive response with temperature and strain rate.

## 3 Part II: Mesoscale Insights into Processing-Nanostructure-Property Relationships

### 3.1 Effects of Biphase Interfaces on Texture Evolution

A promising way to make interface-dominant materials is by severe plastic deformation (SPD). Over the years, many SPD techniques, such as accumulative roll bonding (ARB), accumulative wire drawing and bundling, and equal channel angular extrusion, have been applied to two-phase metals, such as Ag-Fe, Cu-Fe, Ag-Ni, Zr/Nb, and Ag-Cu (Yasuna et al. 2000; Carpenter et al. 2015a, b; Kikuchi et al. 1997; Nizolek et al. 2015).

Figure 3a illustrates the substantial microstructural evolution of a two-phase material during the ARB process. The ARB process involves a sequence of repeated rolling, cutting, and restacking. At the beginning of the process, the metals are coarse-grained sheets with many crystals spanning the distance between two
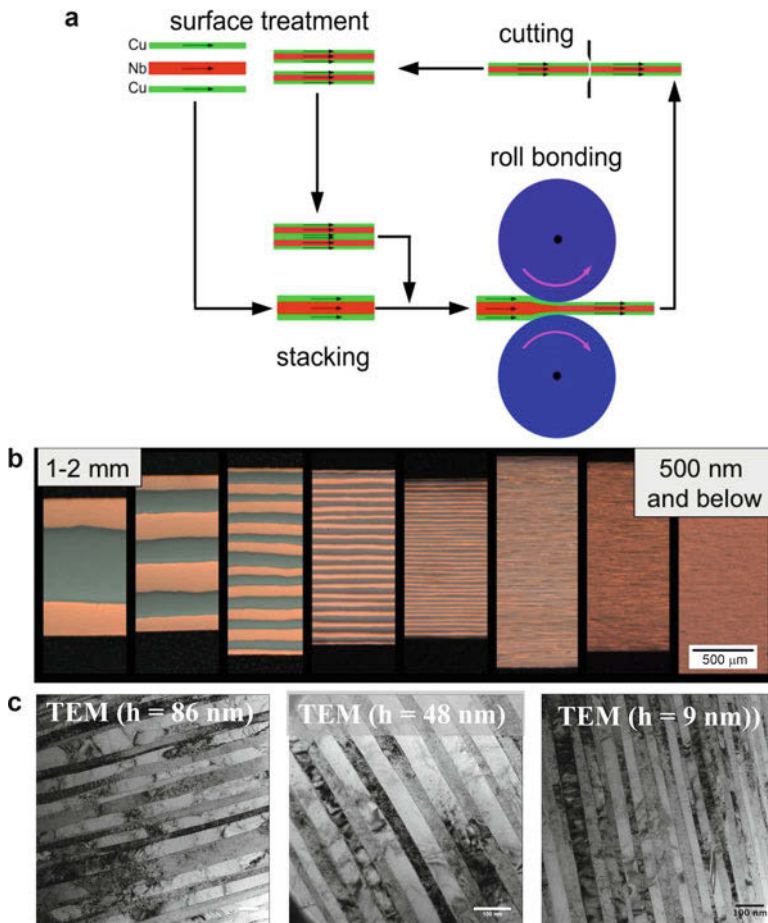
**Fig. 3** (**a**) The severe plastic deformation process, accumulative roll bonding (ARB), used to synthesize. (**b**) Photograph of the Cu/Nb nanolayered composites of controllable layer thickness. (**c**) Transmission electron microcopy (TEM) micrograph displaying the planar Cu/Nb interfaces in an $h = 86$ nm, 48 nm, and 9 nm composites

adjacent Cu/Nb interfaces. With further straining, the spacing h between neighboring bimetal interfaces refines, the bimetal interface density increases, and the grains in each phase become elongated in the rolling direction and finer in the thickness dimension. Figure 3b presents the layer refinement with increasing number of ARB passes for a Cu/Nb composite. Eventually after several passes corresponding to an accumulated strain of ∼10, the layer thickness is refined to submicron dimensions and below, and a single crystal spans the layer thickness (Fig. 3c).

In a few studies, the texture evolution during the ARB refinement process of two-phase laminates (Cu/Nb and Zr/Nb) was studied in order to assess the effect of interfaces on slip or twin activity. Texture evolution during deformation is a result of the activity of slip and twinning in the crystals, and any change in texture indicates
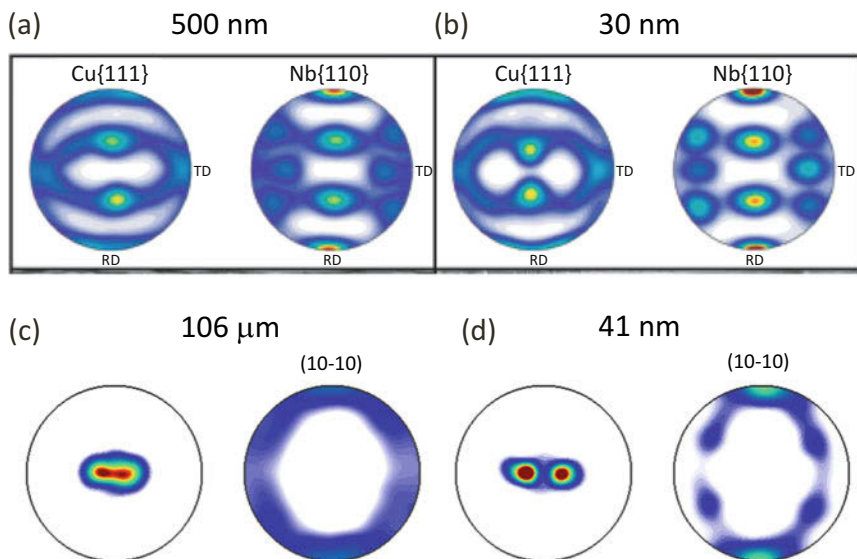
**Fig. 4** Pole figures of the deformation textures measured in the Cu and Nb phase of a layered Cu/Nb composites with individual layer thicknesses of (**a**) $h = 500$ nm and (**b**) $h = 30$ nm. Pole figures of the deformation textures measured in the Zr and Nb phase of a layered Zr/Nb composites with individual layer thicknesses of $h = 106 \mu$ and $h = 41$ nm (Taken with permission from Nizolek et al. (2016) and Carpenter et al. (2015a))

a change in slip or twin activity. These ARB studies used neutron diffraction to measure the texture, as it is a bulk, nondestructive measurement texture technique that enables measurements for layer thicknesses that span from mm to nm. For thick layers (>10 um), the textures of the individual phases corresponded to typical rolling textures, those expected for the phases as if they were rolled alone, as shown in Fig. 4a. However, when the layer thickness reduced below a few microns and the texture strongly deviated from that expected for rolled phases alone. These highly oriented textures of the phases within the nanolayered Cu/Nb composites and Zr/Nb composites made by ARB are shown in the form of pole figures in Fig. 4b. The textures, particularly for the Cu and Zr phases, were strong, containing only a few special rolling components as opposed to the several rolling components commonly expected of rolled single-phase Cu and Zr. Evidently the deformation of the grains and their slip activity must have been augmented by their interface.

These highly oriented textures are associated with the development of a predominant interface with a crystallographic character ranging from $\{1\text{-}12\}<\text{-}111>$FCC$||$:$\{112\}<1\text{-}10>$BCC to $\{\text{-}4\ \text{-}4\ 11\}<11\ 11\ 8>$FCC$||$:$\{112\}<1\text{-}10>$BCC. The crystal orientations correspond to common rolling components in FCC Cu, called C and D, and in BCC Nb called I. In texture nomenclature, they are conventionally indicated as $\{$C:$\{1\text{-}12\}<\text{-}111>$ and D:$\{\text{-}4\ \text{-}4\ 11\}<11\ 11\ 8>$ and I (I:$\{112\}<1\text{-}10>$), given by the crystallographic plane and direction that align with the rolling plane
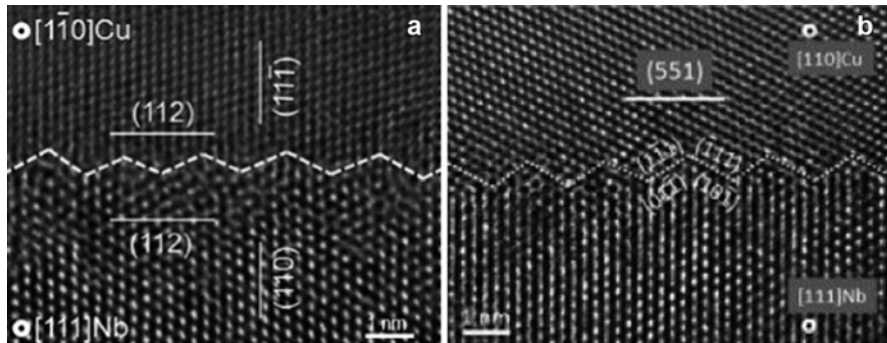
**Fig. 5** Ordered interfaces after extreme strains. High-resolution transmission electron microscopy micrographs of preferred Cu/Nb interfaces: (**a**) {112}<111 >Cu||{112}<110> Nb and (**b**) {551}<10 1 1 > Cu||{112}<110 > Nb. The crystallography of the facet planes is indicated respectively (Taken with permission from Zheng et al. 2013)

and rolling direction. Borrowing this nomenclature, these interfaces were denoted as those ranging from C/I interfaces to D/I interfaces, indicating that the C or D rolling components in Cu are joined to the I rolling component in Nb. The selection of these particular Cu and Nb components is interesting since the many other stable rolling components in Cu and Nb did not appear. Last, these interfaces were observed to be atomically pristine and ordered (see Fig. 5), which is another observation not expected of interfaces formed after severe plastic deformation.

To help understand the texture transition in Cu/Nb, electron backscatter diffraction (EBSD) analysis was used to locate orientations and phases within the microstructure. The EBSD analysis revealed that the transition is coincident with attainment of layers that are spanned by only one grain. In the coarser, micron-layered ($h = 45$ μm, 20 μm, and 8 μm) ARB Cu/Nb composites, several grains still spanned the layers. Within these composites, those Cu grains joined to the bimetal interface, called the interface grains, had predominantly C and D orientations, and those Nb grains had the I component Fig. 6b. The internal bulk grains away from the interface and bounded only by grain boundaries had a typical rolling texture Fig. 6a. This result indicates that the severe plastic straining process of ARB created preferred C/I to D/I interfaces even in the thicker layered composites.

To date, there is no standard theory to explain the development of a preferred texture in an interface-dominant polycrystal produced via rolling or large-strain metal-forming processes. The texture analyses show that the crystals joined by these interfaces correspond to well-known stable orientations in rolling. These are special orientations in which the crystals can plastically deform with negligible lattice reorientation. Based on this idea, the concept of *orientation-stable interfaces* was proposed; that is, for other deformation modes apart from rolling, the character of orientation-stable interfaces, the crystals on both sides are deforming yet neither is reorienting, so the interface character is preserved. When either or both crystals reorient, then the interface character is not plastically stable.
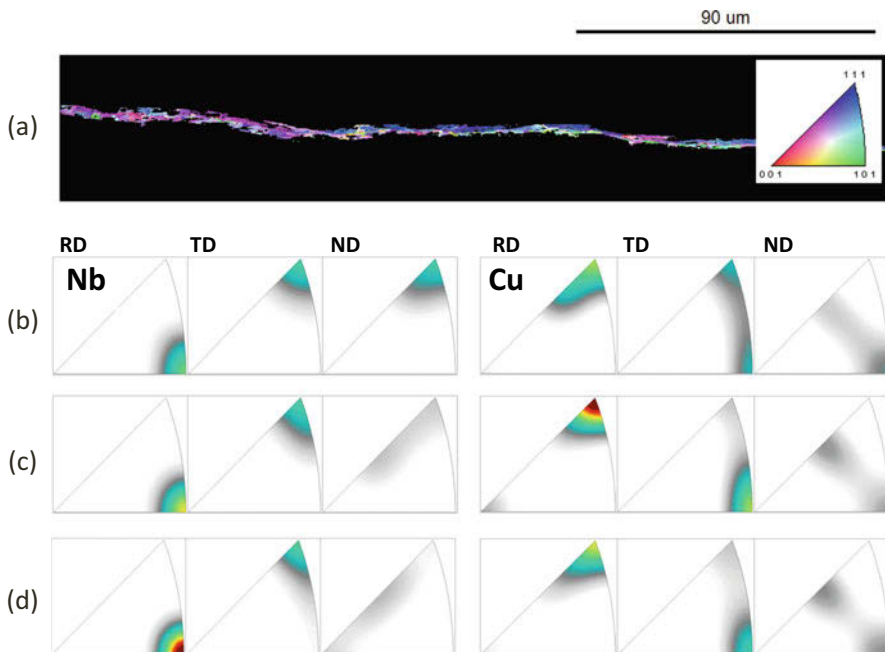
**Fig. 6** (**a**) EBSD showing the texture of interface only grains ($h = 45\,\mu$). (**b**) Inverse pole figures showing the texture of the Nb phase and Cu phase in thick polycrystalline layers. (**c**) Inverse pole figures showing the texture of the interface grains in the Nb phase and interface grains in the Cu phase in thick polycrystalline layers. (**d**) Inverse pole figures showing the texture in the Nb phase and Cu phase in the 200 nm layered composites (Taken with permission from Carpenter et al. 2014)

The concept of mechanical stability for an interface can be tested using CPFE. With this technique, the structures and crystallography of the grains, grain boundaries, and bimetal interfaces can be directly represented in the calculation, and the result is the spatially resolved displacement and orientation gradients at the subgrain scale. Figure 7 shows a few example cross-sections of Cu/Nb microstructures simulated using CPFE. The central grain pair indicated in the figure makes up part of the interface formed between two polycrystalline layers in the simulation volume. It is given a special Cu/Nb orientation relationship and interface plane prior to orientation stability calculations. These deformation simulations account for the differences in elastic anisotropy, viscoplastic flow behavior, and crystallographic slip systems in individual crystals Cu and Nb. Some of the key assumptions made were that the bimetal interfaces and grain boundaries are constrained to maintain compatibility and the rolling deformation is idealized as plane strain compression. The observation of co-deformation and no signs of interface failure support the former assumption. The latter assumption has been shown valid for the material in the center of the sheet away from direct contact with the rolls (Zecevic et al. 2016).

To examine plastic stability from these deformation simulations, the absolute lattice reorientation angle from the original starting orientation is calculated and
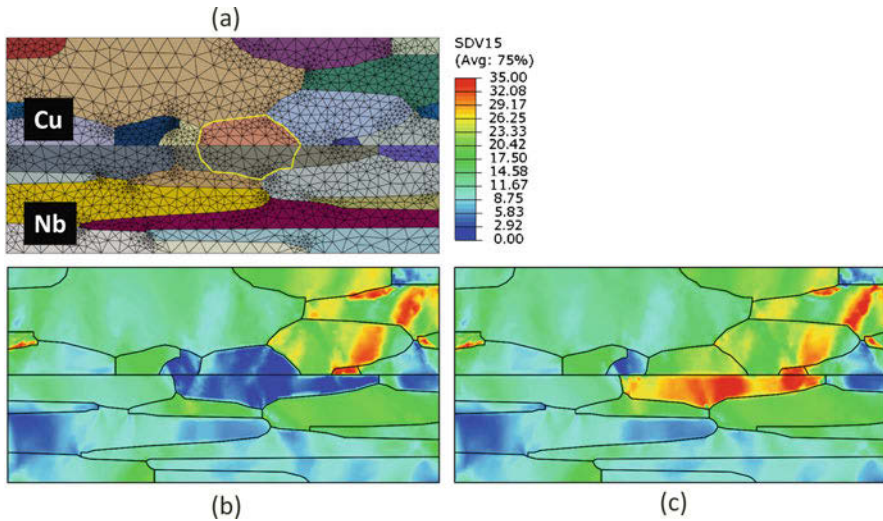
**Fig. 7** Crystal plasticity finite element mesh models of a two-phase layered composite (**a**) model of a grain structure mapped from experimental data and (**b**) reorientation maps of two-phase grain structures deformed in plane strain compression. The top phase is Cu and the bottom phase is Nb. The central embedded bicrystal outlined in yellow has a prescribed starting interface crystallography. In (**a**), the crystallography corresponds to the C/I interface and in (**c**) the KS interface, which is a low-energy interface characteristic of PVD Cu/Nb nanolayered composites. The crystallographic orientations are overlaid onto the original mesh seen in (**a**). The deformed mesh is not shown (Taken with permission from Mayeur et al. 2014)

mapped onto the original (undeformed) microstructure. In these maps, blue means small amounts of lattice reorientation (<6°), whereas yellow to orange correspond to large reorientations (>25°). As an example, Fig. 7b, c compares the reorientation maps for two different interfaces (C/I and Kurdjumov-Sachs (KS) interfaces). The C/I interface maintains good plastic stability in the polycrystalline environment, apart from some variation in misorientation near the grain boundaries. In contrast, the KS interface ({111}<110>Cu||{110}<111>Nb), which develops in the Cu/Nb PVD nanolayered foils, develops relatively large and inhomogeneous reorientation fields. This result suggests that these interfaces, once created, can remain stable, as the layers are refined from microns to nanometers.

When the grains are bounded only by interfaces, such as in the ultrafine grain layers ($h < 2$ μm), the grains become even more constrained and phase stability compromised further. To investigate the plastic stability of the grains in this scenario, Fig. 8 shows the calculation for the same two Cu/Nb interfaces, created by grains lying between two adjacent interfaces, using the same color scheme for the reorientations. For the C/I interface created in ARB, the reorientations are minimal and homogeneous. Again, the analysis predicts that the KS interface is plastically unstable. In fact, the CPFE bicrystal model predicts that the PVD interface reorients toward the C/I interface after large rolling reductions, which is a testament to the stability of the C/I interface.
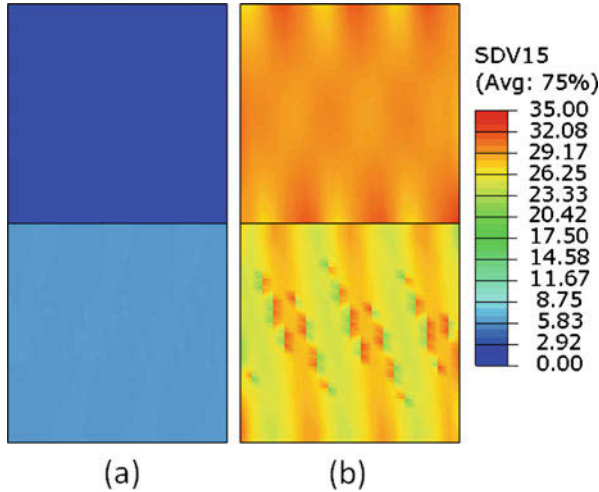
**Fig. 8** Reorientation maps of bicrystals deformed in plane strain compression. The top crystal is Cu and the bottom crystal is Nb. The starting interface orientation relationships for these bicrystals correspond to the (**a**) C/I interface and (**b**) KS interface, which is a low-energy interface characteristic of PVD Cu/Nb nanolayered composites. The crystallographic orientations are overlaid onto the original mesh. The deformed mesh is not shown (Taken with permission from Mayeur et al. 2015)

Taken together, mesoscale modeling using CPFE finds that the compatibility constraint of code-forming layers in two-phase laminate composites can select favored interface crystallographic characters, ones that exhibit orientation stability. These orientation-stable interfaces persist at the interface in coarse layers and remain dominant as layers refine to be one grain thick.

## 3.2    Incorporating Slip/Interface Interactions into Crystal Plasticity Models

An important boundary/dislocation process is slip transmission. It is a mechanism for transferring slip from the crystal on one side of the interface to the crystal on the other side, thereby facilitating plastic deformation in polycrystalline materials. Some grain boundaries or interfaces are more efficient than others at transferring slip, and in this way, the type of interface that exists in the material can profoundly affect plastic flow and ductility.

Given its importance, many experimental studies on polycrystalline metals have studied slip transmission across grain boundaries and bimetal interfaces (Wang et al. 2008, 2012; Beyerlein et al. 2015; Hunter et al. 2018). The general consensus is that in most cases, slip transmission across a boundary is difficult. Thus, dislocations are not expected to freely pass from one crystal to another across an interface via any pathway.

Considering a two-phase crystalline material, slip transmission across a bimetal interface is possible at the head of a pileup that has developed in a coarse-layered composite, in which pileups tend to generate. However, if the layers are too fine to support pileups, slip transmission is expected to occur at the very fine scales (<10 nm) when the stresses become sufficiently high (Misra et al. 2005). In fact in the finest of nanolayered composites, slip transmission can, therefore, play a role in determining the strength of the composite.

There are various scenarios in which to describe slip transfer across an interface. In coarse crystals, dislocations belonging to one slip system α in crystal A can pile up at an interface and invoke nucleation of dislocations on another slip system β on the other side in crystal B. Alternatively, if the crystal is too fine to develop a pileup of dislocations, the stress field of an approaching α dislocation in crystal A can promote nucleation of a β dislocation on the other side in crystal B. In a related scenario, the impinging α dislocation in crystal A is first absorbed into the interface, and then a β dislocation may emit into crystal B in order to lower interface energy. The energetics depends largely on how the interface restructures the core of the absorbed dislocation.

Two factors determine whether slip transmission occurs when a dislocation intersects with an interface. One factor is geometric, having well-aligned slip planes and slip vectors, and the other factor mechanical, possessing driving forces directed into and out of the interface. The number of such favorable pathways across an interface would depend on the crystallographic character (five-parameter description including the orientation relationship and interface plane) of the interface and its orientation with respect to the loading state. An interface character that provides a significant barrier to slip has little to no efficient transmission pathways. In contrast, an interface character that readily transfers slip has many favorable transmission pathways and provides little resistance to slip.

An interface can evolve with deformation to be either more or less favorable for slip transfer. A 3D full-field, spatially resolved CP technique is well suited for modeling microstructure and texture evolution in two-phase materials with bimetal interfaces, but many conventional versions do not take into account the effect that interface/slip interactions can play in microstructural evolution.

Recently, a first attempt was made to examine the effect of slip transmission on slip activity in the adjoining crystals and in turn the evolution of interface character (Mayeur et al. 2015). The authors incorporated into the two-phase CPFE model, a combined geometric and local stress criterion for slip transmission across an interface. It was based solely on the degree of alignment of their slip planes and slip directions on either side of the interface and whether or not the local mechanical driving forces supported the flux into and out of the interface (Misra et al. 2004).

The CPFE model with slip transmission was applied to study the texture evolution in Cu/Nb nanolayered composites during rolling. It was seen in studies of texture evolution in processed nanolayers that at the finest scales ($h = 20$ nm and below), texture experienced a final transition (Zheng et al. 2013). In the Cu/Nb composites, a particularly odd interface of G/I with a {515}<1 -10 1>|| {1-2-1}<101> interface character had formed and become orientationally stable,

**Fig. 9** Reorientation maps of bicrystals deformed in plane strain compression and with slip activity controlled by slip transmission. The top crystal is Cu and the bottom crystal is Nb. The starting interface orientation relationships for these bicrystals correspond to the (**a**) C/I interface and (**b**) G/I interface. The crystallographic orientations are overlaid onto the original mesh. The deformed mesh is not shown. The legend is the same one as that shown in Fig. 8
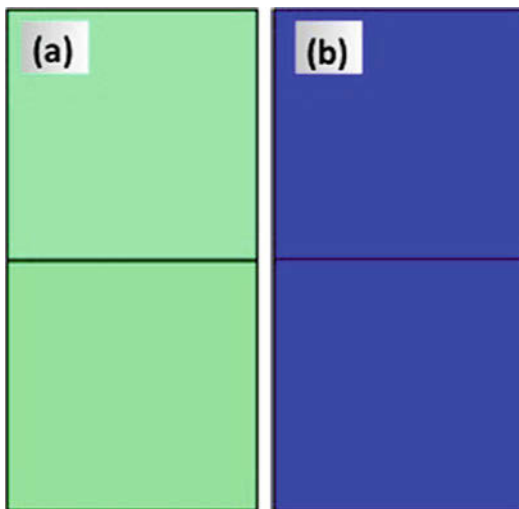
while the previously studied C/I interface and D/I interfaces orientationally unstable (Zheng et al. 2013). It was anticipated that the change in slip activity in the Cu and Nb phase may be due a transition from slip within the crystals to a prevalence of slip transmission across the crystals as the layers became fine.

Figure 9 shows reorientation maps for the C/I and G/I interface with slip transmission-dominated slip as calculated from the model. To see the effects of slip transmission, these maps can be directly compared to maps in Fig. 8 calculated with the standard CPFE and ordinary slip. In all cases, it was found that slip transmission across the interface significantly altered slip activity. The alteration causes the C/I interface (and D/I interface as well) to become unstable (reorientations greater than 12°). In contrast, the G/I interface maintained its character in rolling. The results show that the G/I interface is stable with and without slip transmission-dominated slip, even though the slip activity in these two cases is not the same.

To summarize, the mesoscale model analysis indicates that the transition in texture evolution to a particular stable interface was the result of slip transmission across the interface. Slip transmission can predominant at the finest scales, and when it does, it can affect texture evolution, producing crystalline orientations unlike those expected when conventional slip processes are prevalent.

The slip transmission-dominated calculations assume that only slip systems belonging to favorable slip transmission pathways are active. In actuality, the constraint may be less severe as the interface can supply dislocations by other means (e.g., direct nucleation at triple junctions). Coupling all relevant interface-driven plasticity phenomena, from dislocation nucleation, annihilation, transfer, and storage, has yet to be done. It would be a challenge especially well suited for mesoscale modeling approaches to take.

## 3.3 Nanocrystalline Materials

Nanocrystalline (NC) metals are another example of a boundary-dominant metal. NC are polycrystalline materials that have grain sizes smaller than 100 nm and for decades have attracted a lot of attention because of their ultrahigh strength and hardness, being several times that of their coarse-grained counterparts (Dalla Torre et al. 2002; Ebrahimi et al. 1999; Hughes et al. 1986; Schuh et al. 2002; Schwaiger et al. 2003; El-Sherik et al. 1992; Xiao et al. 2001; Godon et al. 2010; Champion et al. 2001; Chen et al. 2006; Das et al. 2006; Khan et al. 2008; Shen et al. 1995; Youngdahl et al. 1997; Youssef et al. 2004; Nizolek et al. 2016). These experimental reports have shown that the yield strength of NC metals increases with decreasing grain size $D$ for $D$ larger than a critical size, e.g., 10 nm. Most often this size effect on strength closely follows an empirical Hall-Petch (H-P) relationship, in which the yield strength is inversely proportional to the square root of D. Figure 10 shows a compilation of strength data from many studies on NC Cu and NC Ni on a H-P plot. The variability among the studies originates from different measures of strength used, manufacturing methods, and other microstructural features, such as texture or variability in grain size.

NC metals are exceptionally strong because they contain an unusually high density of grain boundaries (GBs), which act as sources and sinks for dislocations and significantly modify dislocation motion. In situ transmission electron microscopy (TEM) and molecular dynamics (MD) studies of deformed nanocrystalline materials have provided insight into how dislocations move within the nanograins (Kumar et al. 2003a, b; Li et al. 2009). They suggest that the prevailing dislocation
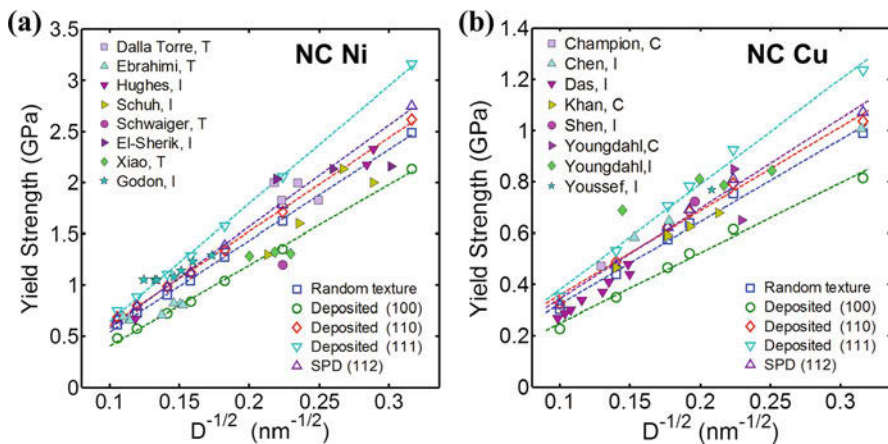


**Fig. 10** (**a**) Comparison of predicted yield strengths with experimental results on NC Ni from various groups. (**b**) Comparison of predicted yield strengths with experimental results on NC Cu from various groups. The dashed line represents the Hall-Petch fit to the calculated yield strength. Experimental data were obtained by nanoindentation marked with an "I," compression marked with a "C" for tension marked with a "T." (In nanoindentation tests, the yield strength was approximated as hardness divided by 2.7)

motions are very different than those in coarse-grained materials. First, dislocation-dislocation interactions and dislocation accumulation seldom happen within nano-sized grains. Second, the plastic deformation proceeds by discrete and separated dislocation slip events, in which a dislocation originates at a grain boundary, glides across the grain unhindered, and recovers at an opposing grain boundary.

NC materials can be made in different ways, and each can lead to different textures. Electro- or physical vapor deposition results in a highly textured NC material, where one set of planes is aligned along one direction and the remaining axes are otherwise randomly oriented in the other two orthogonal directions (Meyers et al. 2006). Plastic deformation (by shearing, rolling, or extrusion) tends to produce even stronger textures than deposition, where the interface planes correspond to the rolling (or shearing) plane and the two in-plane directions are highly aligned with the rolling (or shearing) direction (Nizolek et al. 2014). On the other hand, other methods for making NC metals, such as ball-milling and consolidation processes, produce nearly randomly oriented textures (otherwise known as "no texture") (Khan et al. 2008).

Understanding the degree of coupling between texture and nanograin deformation is important for interpreting reports on strength-grain size scalings in NC materials. In most strength studies performed on the same NC material, such as those in Fig. 10, samples with different grain sizes do not necessarily have the same texture, or quite often, samples from different studies with the same average grain size have different textures due to processing differences. A few recent studies, dedicated to analyzing both texture and size scaling in NC materials, have reported a strong texture effect. Dalla Torre et al. (2002) investigated two commercial NC Ni samples with the same grain size ($\sim$20 nm) and clearly demonstrated that the strong (100) and weak (100) initial textures can result in non-negligible differences in yield strength, ultimate tensile strength, and plastic strain. Godon et al. (2010) studied the effect of crystallographic texture on the relationship between grain size and flow stress in NC Ni. Their experimental results showed three distinct regimes in the Hall-Petch plot, corresponding to samples with a (100) texture, (110) texture, and random texture. They concluded that differences in Hall-Petch slope resulted from differences in the deformation mechanisms induced by texture.

The physical picture provided by experiment and atomic simulation implies that the dislocation processes in NC are closely tied to the availability of dislocation sources in the GBs and orientation of the dislocation glide planes with respect to the grain boundaries. On this basis, texture effects ought to be stronger in NC metals than they are in coarse-grained (CG) metals. CP-based techniques are well suited for assessing coupled texture and grain size effects in NC materials and how they are coupled to determine material strength. As discussed in Sect. 2.4, the common CRSS models for activating slip with crystal plasticity-based models, however, apply to coarse-grained materials, and these models assume that slip initiates and propagates homogeneously within the grains, without resolving the individual contributions of the dislocations. The mesoscale modeling challenge, therefore, lies in how to represent the dislocation glide processes in NC materials so that texture/grain size influences can be properly replicated in simulation.
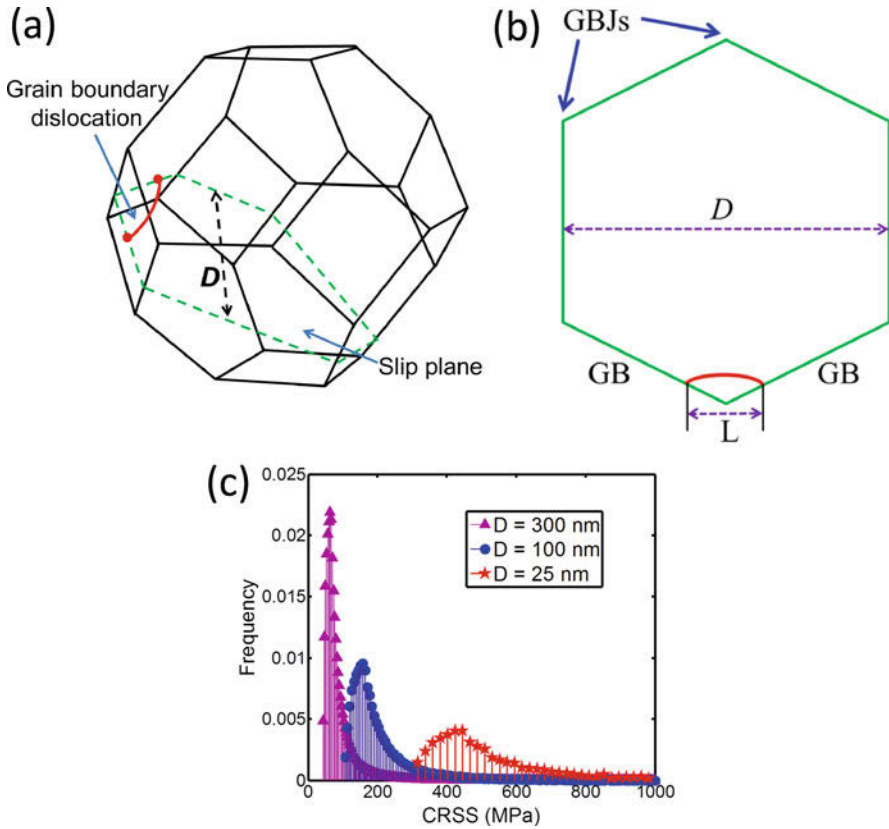
**Fig. 11** (**a**) Schematic of a dislocation source emanating from a grain boundary triple junction in a nanograin. The grain is embedded in a polycrystal, and the grain boundary facets that it makes with the neighboring grains give it a shape of a tetrakaidecahedron. (**b**) Schematic of a dislocation source (red) at a grain boundary triple junction. The length L is the dislocation source length. The shape of the grains is assumed to take on a regular hexagon. (**c**) Comparison of the probability density distribution of CRSS in NC Cu with grains sizes of 20 nm, 100 nm, and 300 nm (Taken with permission from Yuan et al. 2015)

Recently Yuan et al. (2015) developed a discrete-slip-crystal-plasticity-based CPFE model in which slip occurs in discrete slip events exclusively by individual dislocations emitted statistically from the grain boundaries. Within the model, the shape of the nanograins is assumed to be tetrakaidecahedron, that is, a truncated octahedron, such as the one shown in Fig. 11a. In this case, some dislocation glide planes assume a hexagon shape, while others would adopt a pentagon shape. Within this plane, dislocation sources are present in the grain boundary as double-pinned dislocation segments of length $L$ emanating from the triple junctions as illustrated in Fig. 11b.

Given the length $L$, of the double-pinned GB source, the characteristic stress to bow out the segment to an unstable configuration is given by Foreman's formula (Foreman 1967):

$$\tau_{CRSS} = \frac{\mu b}{2\pi L} \log\left(\frac{L}{r_0}\right) \tag{14}$$

where $\mu$ is the effective shear moduli of the crystal and $r_0$ is the core radius of the dislocation. Once emitted from a GB, the dislocation glides unhindered to the opposing grain boundary producing a shear strain given by

$$\gamma_{discrete} = c\frac{b}{D} \tag{15}$$

where $c$ is a scaling parameter independent of grain size ($c = 1.2$ in their work), $b$ is the magnitude of the Burgers vector, and $D$ is grain size.

These source lengths $L$ are statistically distributed where $L$ can be no shorter than a dislocation core $L_0$ and no longer than $D$ the diameter of a grain. Combined with the expression for the activation stress in Eq. (14) leads to a distribution for CRSS. Figure 11c is an example of a CRSS distribution for three different grain sizes, when $L$ is assumed to be equal likely to lie between $L_0$ and $D$. As shown, by virtue of statistical source lengths from grain boundaries, dislocation activation becomes harder (the statistical mean increases) and more variable (the statistical dispersion increases) as $D$ decreases.

Yuan et al. (2015, 2016a) applied the discrete-slip-CPFE model to study the combined effects of grain size and texture on the strength of NC materials. As common in CPFE simulations for coarse-grained materials, glide on multiple planes is permitted and dictated by the current stress state in the grain, and its current orientation and grain-grain interactions are taken into account. However, in the discrete-slip model for nanomaterials, during deformation, for each GB junction, a GB source is randomly assigned a $\tau_{CRSS}$ from the corresponding $\tau_{CRSS}$ distribution (Fig. 11c). Without loss of generality, the statistical CRSS model is applied to the case in which the GB sources are inexhaustible. After emission of a dislocation on slip system $s$ from a GB source, a new GB source must take its place in order to emit another dislocation on the same slip system. After emission, a new $\tau_{CRSS}$ is reassigned for a potential slip event. As validation, with only the core radius and strain rate sensitivity used as material input, the model is able to reproduce the entire stress-strain curve measured for a few grain sizes of NC Cu, as shown in Fig. 12.

The discrete-slip CPFE simulation model was employed to understand the effects of grain size and texture on NC material strength in simulation, the deformation was applied in one of three directions at a strain rate typical of most laboratory studies ($10^{-4}$/s). Additionally, as mentioned the reported NC studies (see Fig. 10), the initial textures very likely vary due to differences in processing. To account for processing-related texture effects, these simulations began with different types of initial textures, either random or highly oriented, deposited textures. The

**Fig. 12** Comparison of experimental and calculated stress-strain curves for texture-free NC Cu with various grain sizes. (Experiment data are taken from Khan et al. 2008) (Taken with permission from Yuan et al. 2015)
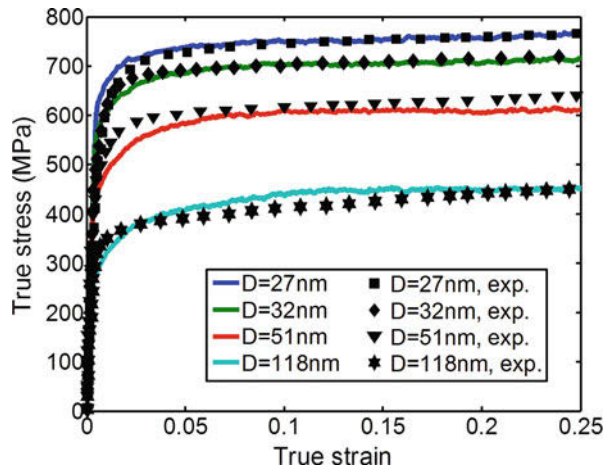
orientations in the model material for the former case of a random texture were numerically generated. For the latter case of a highly oriented texture, a set of deposited {hkl} textures, where the {hkl} planes are aligned along the *z*-direction and nearly isotropic in the *x-y* plane, were created. These include deposited {001}, {110}, and {111} textures.

Figure 10 presents, along with the strength data from several studies, the model results for the 0.2% yield stress taken from the calculated stress-strain curves for the different loading directions and initial textures (Yuan et al. 2016b). The results bound the corresponding experimental data. The important implication is that initial texture differences in NC samples alone can result in a significant dispersion in strength.

The straight line that the predictions form on this plot for each initial texture shows an emergence of a $D^{-1/2}$ Hall-Petch scaling in the calculated yield strength. The Hall-Petch relationship prevails for all initial textures and materials studied, with the only difference being a change in the Hall-Petch slope. Further, the model finds that the more intense the texture (more preferred orientations), the larger the Hall-Petch slope, indicating greater strength sensitivity to grain size reduction.

A commonly used metric for relating texture intensity to the anisotropy in strength is the Taylor factor $m_T$. It is defined as the inverse of the average of the (five) largest Schmid factors in each grain, which is then averaged over all grains. It is thus a purely geometric factor that depends on loading direction and initial texture. For instance, for the z-direction for the {111} texture, $m_T^z = 3.44$. The model results suggest that the H-P coefficient can be expressed as a product of a material-dependent parameter and a geometric one, i.e., $\alpha m_T$, giving the following Hall-Petch law

$$\sigma \propto \alpha m_T D^{-1/2} \quad (D < 100 \text{ nm}) \tag{16}$$

For Ni and Cu, $\alpha$ is, respectively, 3.22 GPa• nm$^{1/2}$ and 1.14 GPa• nm$^{1/2}$. If for instance, the plastic anisotropy measure is chosen to equal the difference between the z-direction and x-direction yield strength, $\psi = |\sigma_z - \sigma_x|$, then from Eq. (16), the following expression for the H-P coefficient can be written

$$\psi = \alpha \Delta m_T D^{-1/2} \tag{17}$$

where $\Delta m_T = \left| m_T^z - m_T^x \right|$ . For a purely random texture (no texture), then $\Delta m_T$ is zero and the anisotropy is zero. Otherwise, for a material with a texture, $\Delta m_T$ is nonzero. Equation (17) indicates that when the material possesses a strong texture, the plastic anisotropy associated with a textured material can increase as the grain size reduces. This outcome is a demonstration of a mesoscopic effect, involving an interaction of two mesoscale microstructural features.

## 3.4    Nanotwinned Materials

Nanotwinned (NT) metals are another class of boundary-dominant materials, like NC metals. Their structure is comprised of a large density of twin boundaries, spaced usually 5–80 nm apart. These metals have gained much attention due to reports of excellent strength and hardnesses, four to six times higher than those of their bulk counterparts, while exhibiting reasonable ductility (Lu et al. 2004, 2009a; Beyerlein et al. 2014). They are also found to be harder than NC or ultrafine-grained forms of the same material and possess outstanding thermal stability (Zhang et al. 2005, 2007; Anderoglu et al. 2008), electrical conductivity (Chen et al. 2007), and fatigue resistance (Pan et al. 2013; Pan and Lu 2014).

Most NT materials have an FCC crystal structure, such as Cu and Ag, and are formed by deposition techniques (e.g., electrodeposition, magnetron sputtering) (Shen et al. 2005; You et al. 2011). The nanotwinned microstructure is comprised of relatively large diameter $D$ grains (1–5 μm) that are laminated with twins with an average nanoscale thickness of $\lambda$ (5–80 nm) Fig. 13). There are two types of NT materials: columnar and equiaxed. In equiaxed NTs, the grains are randomly oriented, while in columnar-grained NTs, the grains form as columns aligned along the NT growth direction. They consequently have very different textures. The TBs and {111} poles are randomly oriented in the equiaxed NTs, while they are highly oriented in the columnar NT, being narrowly dispersed about the NT growth direction.

It is generally known that fine nanoscale dimensions of the twin lamella give the NT its high strength. NT materials deform by the glide of (partial or full) dislocations within the twin lamellae on {111}<110> slip systems or on the TB planes. The formation and glide processes are unlike those in coarser twinned or coarse-grained materials. Many MD simulations on deformed NTs (Dao et al. 2006; Jérusalem et al. 2008; Mirkhani and Joshi 2011, 2014; Xie et al. 2015; Zhu et al. 2011, 2015) have revealed that GBs and the GBJs, where GBs meet, are
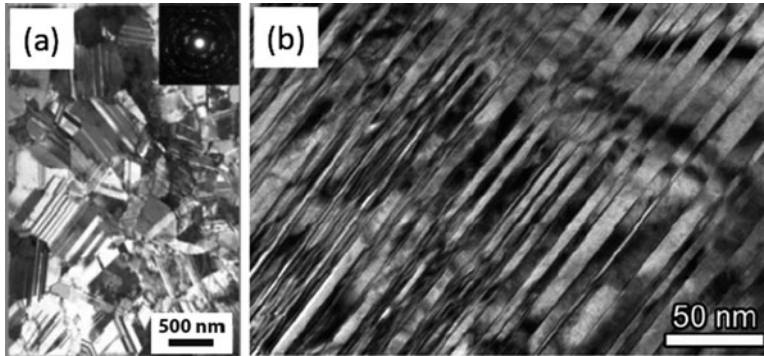
**Fig. 13** (**a**) TEM images of an equiaxed Cu NT and (**b**) a columnar Cu NT (Taken with permission from Lu et al. (2009b) and Zhu et al. (2015))
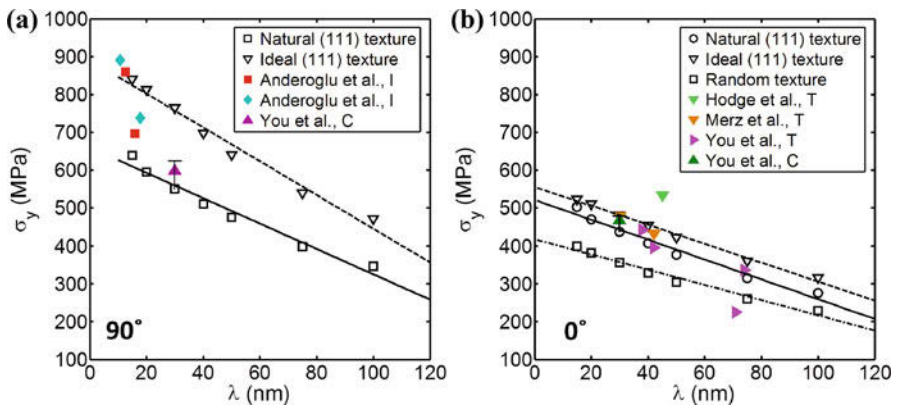


**Fig. 14** Comparison of experimental and calculated yield strengthens for columnar-grained NT Cu (**a**) under 90° compression and (**b**) under 0° compression. The error bars on the calculated strengths for $\lambda = 30$ nm cover results for $D = 500$ nm to 3000 nm. Experimental data were taken from various studies on Cu NT as indicated by solid symbols (Taken with permission from Yuan et al. 2016)

the favorable dislocation nucleation sites. The motion of the dislocations that must glide between adjacent TBs is constrained and the nanotwin thickness $\lambda$ limits the distances traveled. As with NC metals, the formation of dislocation pileups in such a confined space is unlikely.

These mechanisms of deformation indicate that the strength of NTs would depend on the TB spacing (or equivalently, the nanotwin thickness $\lambda$) as well as the as-processed texture (whether {111}||NT or random). Effects of $\lambda$ size and texture are not explicitly examined in many studies. A compendium of NT strengths, measured by indention, compression, or tension, from various studies, is given in Fig. 14. While it appears that strength increases with decreasing $\lambda$, the data are not sufficient to extract a well-defined $\lambda$ scaling. Strengths tested normal (90°) to

**Fig. 15** Experimental results from You et al. and calculated stress-strain curves for columnar-grained NT Cu with twin thickness $\lambda = 30$ nm and different grain sizes $D$ (Taken with permission from Yuan et al. 2016)
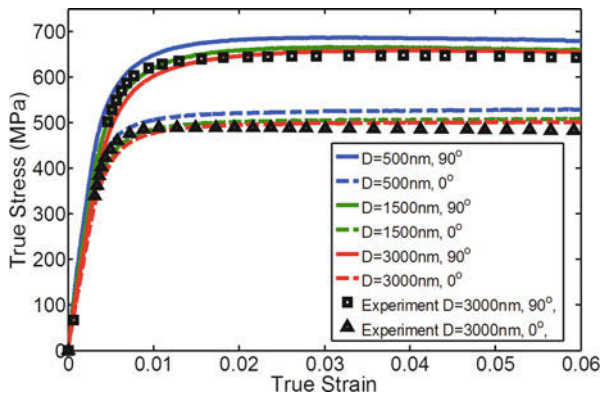
the TBs or parallel (0°) to the TBs are shown in separate plots, since the strengths in these two directions could be dissimilar. In one study, performed by You et al. (2013), the yield strength of columnar-grained NT Cu was shown to be highly anisotropic. When the loading orientation is perpendicular to the TBs (90°), the yield strength is 600 MPa, 25% higher than 0° (see Fig. 15). In contrast, equiaxial-grained NT Cu exhibits near-plastic isotropy because of the randomly oriented twin layers (Lu et al. 2009b).

Recently Yuan et al. (2016b) advanced a mesoscale CP technique to examine the effect of NT $D$ and $\lambda$ and initial texture on NT deformation response, strength, and strain hardening and to investigate the underlying mechanisms. CP, in general, permits the crystalline nanostructured to be modeled closely to the experimental one. At the highest scale is the polycrystal, an aggregate of grains of diameter $D$. Every grain is laminated by multiple NTs, each separated by a (111) twin plane and possessing the FCC twin-matrix orientation relationship. Among the grains, the NT (111) twin planes differ in orientation with respect to the sample, according to a specified texture, e.g., randomly in the case of no texture or highly aligned with the growth direction in the case of columnar NTs. Figure 16 shows a CP model for a columnar NT.

As is common with CP models, at the finest scale, within an individual lamella, plastic deformation is mediated by dislocation motion on crystallographic planes. For FCC crystal structure common of these NT, there are 12 possible slip systems in FCC crystals. With respect to the {111} TBs, there are three slip systems that lie parallel to the TB plane, denoted here as *TB-parallel slip*, and nine slip systems that lie inclined to them, called *TB-inclined slip*.

In many CP techniques, slip on these planes is assumed to activate homogeneously in the crystal and the amount of slip incurred after activation is not explicitly dependent on lamellae size. The slip model used in Yuan et al.'s CP technique (Yuan et al. 2015) (see Sect. 3.2) was modified to account for discrete slip events, where dislocations are emitted from grain boundaries in NT materials and both $\lambda$ and D affect the amount of slip.
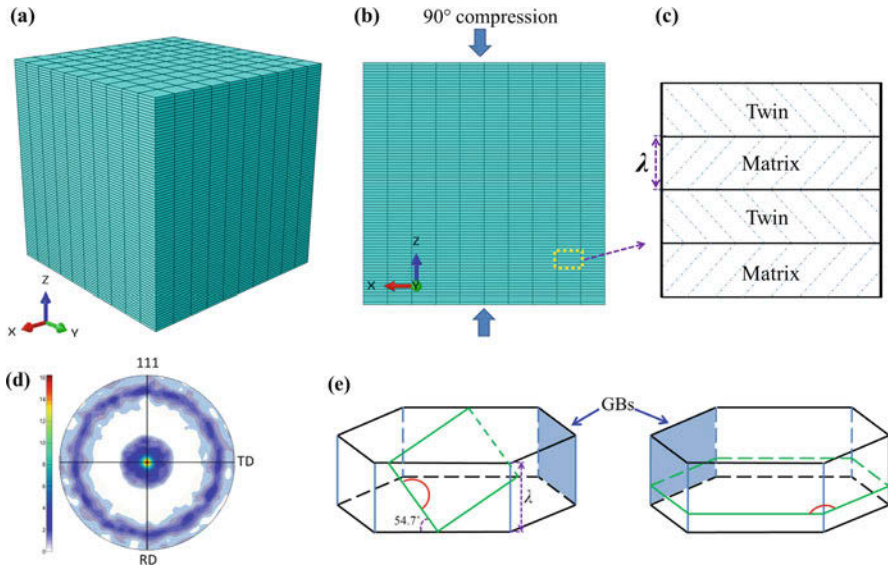
**Fig. 16** (**a**) Schematic of a NT Cu sample with columnar grains. (**b**) Plane view of the NT Cu sample. The compression perpendicular to the columnar axis is defined as 90° compression. (**c**) A portion of a single columnar grain including twin/matrix structure. (**d**) {111} pole figure of the initial texture of the columnar-grained NT Cu. This texture is used to initialize the model microstructure prior to application of the load. (**e**) Schematic of a twin lamella in one grain with a dislocation nucleating from GB on an inclined slip plane (left) and from a GBJ on a parallel slip plane (right) (Taken with permission from Yuan et al. 2016)

Figure 16 shows the model geometry for the case of a columnar NT and, in more detail, a schematic of an individual twin lamella that is $\lambda$ thick. As shown, the cross-section of a twin lamella is modeled as a hexagon. The grain size corresponds to the shortest distance $D$ between two opposing facets of the hexagon corresponding to the grain size $D$. The six sides of the twin lamella are grain boundary (GB) planes. Each lamella has six grain boundary junctions (GBJs), where two neighboring grain boundaries meet.

As mentioned earlier, both GBs and GBJ contain sources for dislocations, which are double-pinned source length lying in the planes or across the junctions. Using the twin lamellae model in Fig. 16, as a GB source, source length $L$ must lie along the intersection line between the slip plane and the grain boundary. These GB sources can supply dislocations on all 12 planes – the 3 TB-parallel slip systems and 9 TB-inclined slip systems. For the TB-parallel planes, the GB source lengths can physically range from $L_0$ to $D$. For the TB-inclined planes, the GB/slip plane intersection line is bound between the two adjacent coherent twin boundaries $\lambda$, and thus, the possible source lengths range from $L_0$ to $\lambda/\sin\theta$, where $\theta$ is the angle between the inclined slip plane and twin plane. For the GBJ sources, the line where two adjacent grain boundaries of the same grain meet, as shown in Fig. 16e. The line does not intersect any TB-inclined plane and therefore are geometrically not

permitted to emit dislocations on them. The GBJ source line does intersect one of the TB-parallel planes. Since each end of the GBJ source length $L$ can lie on one side of the junction, the length $L$ of the GBJ sources can range from $L_0$ to $D$. Finally, the stress to emit the source of $L$ is given in Eq. (14). Statistical variation in $L$ leads to a statistical distribution of activation stresses.

Once emitted from the GB or GBJ source, $D$ or $\lambda$ determines the amount of slip the dislocation delivers. The TB-parallel systems provide a shear of $b/\lambda$, and the TB-inclined systems can shear by $b/D$, where $b$ is the magnitude of Burgers vector for the material (e.g., for Cu, $b = 0.362$ nm).

With the above formulation, the CP technique considers two NT length scales, twin thickness $\lambda$ and grain size $D$, and their size effects on two phenomena, dislocation emission, and the amount of slip per emission. For demonstration the simulations considered the 0.001/s room temperature, strain rate tests by You et al. (2013) on Cu. The model NT Cu initial texture was taken from their study and corresponds to one naturally found in most as-deposited NT materials, wherein the {111} poles of the twin lamellae are dispersed with in 25° about one axis (denoted as the natural {111} texture). Figure 15 compares the model with the stress-strain curves from You et al. (2013) for the 0° and 90° cases. It is worth noting that apart from the nanostructure length scales and textures, the only parameters that were provided to the model were the strain rate and power-law exponent and for Cu, the moduli, and Burgers vector. The grain diameter and $\lambda$ in the model corresponded to those in the actual sample $D = 3000$ nm and $\lambda = 30$ nm. As shown, other grain diameters ($D = 500$ nm and 1500 nm) were also tested to assess the effects of $D$ in this case. The model suggests little dependence on $D$, which can be anticipated, since comparatively the very fine nanotwin thickness $\lambda$ would have more impact on dislocation emission.

To further investigate processing texture effects, the calculation for NT material with $\lambda = 30$ nm and $D = 3000$ nm (in Fig. 15) was repeated for a weaker texture, i.e., no starting texture (random) or a stronger one, an ideal (111) starting texture. Without an initial texture (i.e., random texture), the difference between the strengths for the 90° and 0° was negligible. This result is consistent with reports on equiaxial-grained NT Cu, which exhibits near-plastic isotropy because of the randomly oriented twin layers (Lu et al. 2009b). In contrast, with an ideal (111) initial texture, the plastic anisotropy is significant, with the 90° test being the stronger of the two, as seen in the calculations with the natural, in experimentally measured texture, and in the experimental mechanical tests. The results find that texture causes the 90° test to have the stronger flow stress than the 0° test.

Based on the foregoing results, it would be expected that the same anisotropy prevails for all values of $D$ and $\lambda$ provided the texture does not change. The effect of NT $\lambda$ was examined with the model for the three types of textures: the random texture, the experimental (natural) texture Fig. 16d, and the ideal (111) texture. The comparison made in Fig. 14 finds that these three sets of calculations collectively span the values of NT strengths reported in the literature, which likewise are derived from samples bearing one of these types of initial textures.

Several interesting implications can be made from the calculations. First, the equiaxed NTs are the weakest and those with the ideal {111} texture NTs are

stronger. The model results suggest that for all initial texture types, the yield strength scales inversely with λ. Yet a cross-effect arises since texture is observed to affect the sensitivity of strength to reductions in λ. The size effect in λ is seen to intensify as the texture sharpens from the no-texture case (weakest effect of size on yield strength) to the case with an ideal {111} texture (strongest effect of size on yield strength). For instance, with the strong, highly oriented (111) texture, the often measured 90° strength for columnar NTs would increase much faster with decreasing λ than equiaxed NTs. Last, the size effect is anisotropic. The 90° strength has a stronger size scaling with λ than the 0° strength. Consequently, the yield anisotropy for the columnar NT, defined as the difference in yield strength between the 90 and 0° tests, increases as λ decreases. As a consequence, finer twin thicknesses would lead to a higher propensity to shear band.

## 3.5    Nanolayered Thin Films

One popular class of interface-dominant materials are nanolayered composite thin films. These materials have a 2D structure and are comprised of a layered stack of dissimilar metals, wherein the individual layer thickness has nanoscale dimensions (usually $h < 100$ nm), and hence the density of interfaces is unusually high (Beyerlein et al. 2012, 2013a, d, 2015; Monclús et al. 2013). Studies on the structural properties of nanolayered films report exceptional properties compared to those of their constituents or volume average values of their constituents, such as strengths that are over five to ten times higher, hardness values that are several orders of magnitude higher, and greater microstructural stability in harsh environments, such as irradiation, impact, or elevated temperatures (Beyerlein et al. 2013a; Mara et al. 2008; Mara and Beyerlein 2014; Han et al. 2013a; Misra et al. 2004, 2007).

As a result of the deposition technique, the films have a special microstructure that is unlike those found in traditional polycrystalline materials and even the single-phase nanocrystalline materials discussed earlier (Sects. 3.2 and 3.3). They are usually made by deposition techniques, such as physical vapor deposition (PVD), which enable fabrication of multilayers with individual layer thickness of nanoscale dimensions, uniformly across the film (Fig. 17). Typically, one grain spans an individual layer, but in the plane of the layer, the layers are polycrystalline. The interface bears a certain crystallographic character, usually low in interfacial energy, and the crystal orientations are highly aligned, giving the film a strong texture. The interfaces tend to be atomically ordered, comprised of relatively few defects, such as misfit dislocations, compared to grain boundaries (Beyerlein et al. 2013b, c; Wang et al. 2013; Kang et al. 2012).

Like in conventional, coarse-grained metals, plasticity in these biphase nanolayered films occurs by dislocation glide. However, the nanoscale dimensions of the layers are close to the size of dislocations. The exceptional plastic properties of nanolayered composites, over the coarse counterparts, are a result of the frequent and close interactions of the dislocations with the bimetal interfaces.
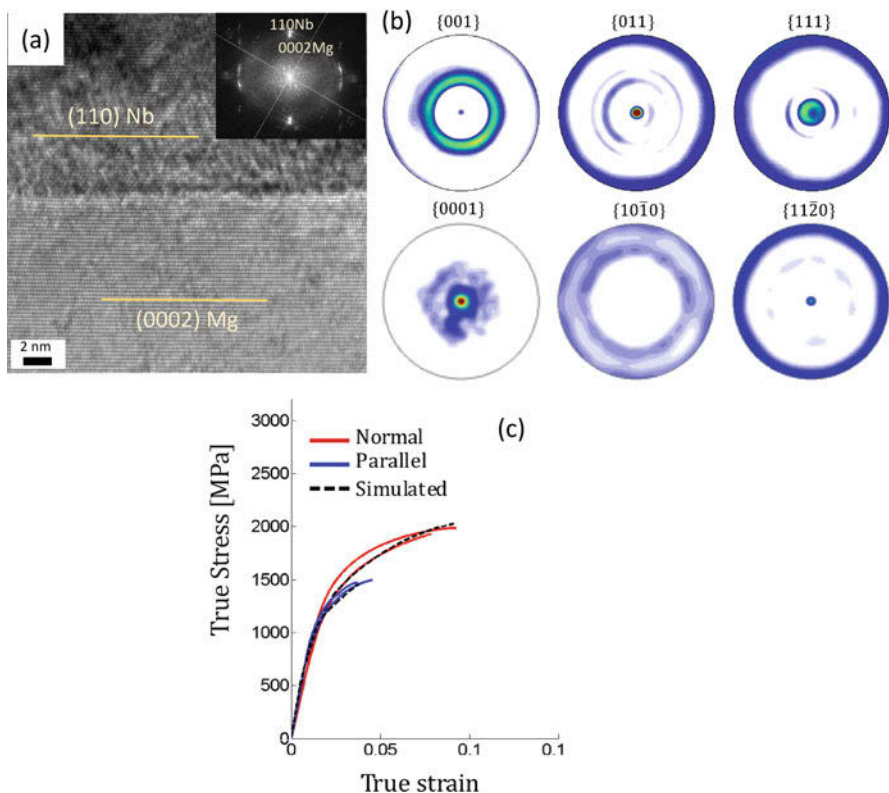
**Fig. 17** (**a**) HR-TEM images taken of the (**a**) 50 nm/50 nm HCP Mg/BCC Nb. (**b**) Initial pole figures of the XRD measured textures of the HCP Mg (50 nm) on the top row and Nb phase in the bottom row. (**c**) Comparison of the predicted and measured stress-strain curves from micropillar compression for the 50/50nm Mg/Nb composite (Taken with permission from Pathak et al. (2017) and Ardeljan et al. (2018))

With crystallographic slip as the main deformation mechanism, the 2D-layered architecture and the strong texture together alone would then be expected to result in substantial anisotropy in strength and other deformation properties. Yet, anisotropy in structure properties is not often studied since testing the thin film samples in different directions is not as straightforward as it is with bulk samples. The few studies in which anisotropy were investigated show considerable anisotropy in yield, strain hardening, and ultimate strength (Nizolek et al. 2016; Pathak et al. 2017). Figure 17 presents results from a recent report on the anisotropy of Mg/Nb nanolayered PVD films. From these films, micron-diameter pillars were fabricated using a focused ion beam (FIB)-based technique and tested in compression to obtain the composite mechanical response. Micropillar compression tests were carried out with the compression axis either (i) normal or (ii) parallel to the Mg/Nb interface planes. In both directions, these materials exhibited outstanding strength, 50% stronger than that of coarse-grained counterparts or a volume average of their

strengths. The composites also exhibited anisotropy in their flow response with the flow stress in the normal direction being 30% higher than in the parallel direction.

In general, crystal plasticity-based techniques are suitable models for calculating the microstructural evolution (texture, layer distortion) during the deformation of layered polycrystalline microstructures. In fact, in recent years, a few applications of CP modeling have led to insights into texture effects. In Al-Fadhalah et al. (2005), a two-phase VPSC model was applied to study texture evolution in rolled PVD Cu/Nb films, showing that texture evolution at a fine scale corresponded to maintenance of the KS orientation relationship at the Cu/Nb interface. In another study, two-phase VPSC successfully showed that the substantial tensile anisotropy in Cu/Nb nanolaminates was the result of the strong texture that developed during processing (Nizolek et al. 2016). A CPFE model for Cu/Nb was used to show that the two common interfaces found in PVD Cu/Nb films did not maintain their crystallography under rolling (Mayeur et al. 2014; Beyerlein and Mayeur 2015). Another Cu/Nb laminate CPFE model was employed to quantify the reductions in micropillar compression strength of Cu/Nb nanolaminates with increases in temperature (Monclús et al. 2013). While calculations from these CP modeling studies explained some trends seen in experiment, predictions of the actual yield strength and flow response were not possible. The constitutive laws traditionally used in CP-based models consider dislocation processes in coarse-grained materials and thus are not appropriate for processes taking place in very fine crystals of nanoscale dimension.

In these nanolayered composites, typically with layer thicknesses $h < 50$ nm, the interfaces are spaced sufficiently close, which is only one order of magnitude than the width of the dislocation core. The interfaces confine the movement of the dislocations, forcing them to thread through the layers, depositing in the interfaces as they pass by. This glide mechanism that has been observed in these nanolaminates materials and is termed confined layer slip (CLS) (Misra et al. 2005; Li et al. 2012).

Recently Ardeljan et al. (2018) proposed a CPFE-based model crystal plasticity model with the $h$-dependent critical resolved shear stresses according to CLS. This CPFE-CLS model employs the CLS model (Misra et al. 2005) to define the CRSS for every slip system operating at a crystalline point. Following the CLS formula, the CRSS for a slip system $s$ is given by

$$\tau_{\text{CLS}}^s = \tau_0^s + A^s \frac{\mu^s b^s}{h^{s\prime}} \ln\left(\frac{\text{ch}^{s\prime}}{b^s}\right) + \frac{f^s b^s}{h^{s\prime}} \tag{18}$$

The first term $\tau_0^s$ represents the sum of a friction stress and resistance from other obstacles, which would not be significantly altered by strain. The second term comes from the CLS mechanism and bears a $\ln(h)/h$ dependence, and results from the resistance encountered by the dislocation as it propagates through the layer. The last term, i.e., $f^s b/h\prime$, results from the deformation of the interface caused by the threading dislocation. Based on previous works that have applied the CLS model to multilayers, the expected order of magnitude for $\tau_0^s$ is 10–100 MPa, for $A^s$ is $10^{-2}$–$10^{-1}$, and for $f^s b$ is 1–3 J/m$^2$.

**Fig. 18** The CPFE model meshes for (**a**) the 50–50 nm Mg/Nb composite. The model contains one layer each of Mg and Nb. Bottom images show zoomed-in view of the microstructure indicating different grain aspect ratio (Taken with permission from Ardeljan et al. 2018)
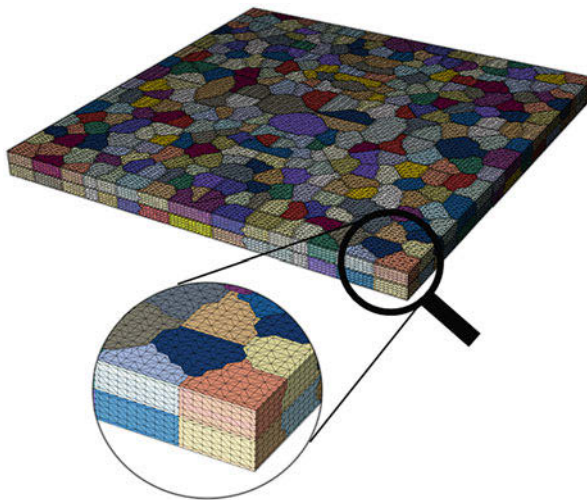
The key length scale introduced in the CLS-CRSS model is $h^{s'}$, the distance along the slip plane of slip system $s$ from one interface to the next. It depends on the orientation of the crystal with respect to the interface normal and the layer thickness $h$. Consequently the CLS-CRSS embodies a dependence on texture, and the CLS-CRSS can evolve with texture through its dependence on $h'$ and with deformation of the layers through its dependence on $h$.

The CPFE-CLS model was applied to understand the mechanisms responsible for the yield, strain hardening, and plastic anisotropy in the Mg/Nb composites shown in Fig. 17c. As is common for the crystal plasticity contribution to the constitutive law, the slip and twin families need to be specified a priori. For HCP Mg, the three main slip modes made available for HCP Mg are basal $\langle a \rangle$ slip, prismatic $\langle a \rangle$ slip, and pyramidal $\langle c + a \rangle$ {1122}⟨1123⟩ slip. For BCC Nb, the two slip modes made available are {110}⟨111⟩ slip and {112}⟨111⟩ slip.

As discussed in Sect. 2.2, part of the effort is the explicit representation of the microstructure in the micropillar sample. Figure 18 shows the CPFE model of the Mg/Nb bilayer. It is divided into two equal sections/volumes that represent a layer of the Mg bonded to a layer of Nb. As in the fabricated film, (1) both layers are polycrystalline in plane, and one grain spans the layer thickness. (2) The grains are nearly equiaxed with grains 50 nm thick and 100 nm wide. (3) The initial Mg and Nb phase textures were produced by assigning orientations randomly selected from their corresponding measured textures. (4) Mg and Nb grain pairs share a common {0001}||{011} interface plane.

Figure 17c compares the experimental and model stress-strain curves, in which good agreement is observed. Figure 19 shows the relative amounts of slip contributed from the different modes according to the model. In the HCP Mg phase, basal slip prevails, as is expected for pure Mg, from the many reports on the deformation of coarse-grained and ultrafine pure Mg-based systems (Zhu et al. 2008).
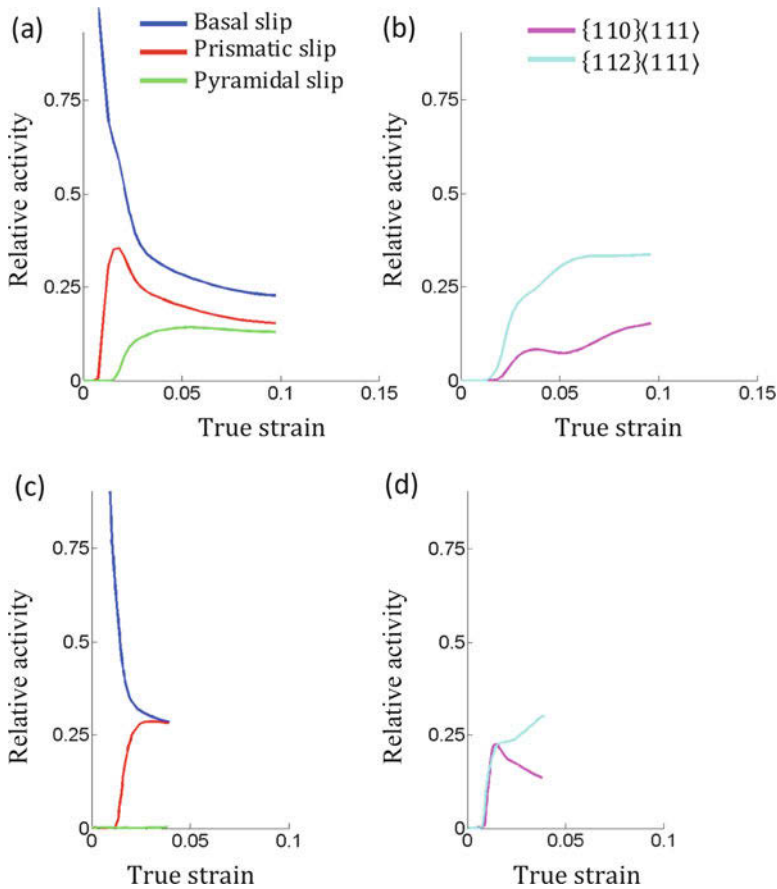
**Fig. 19** Calculated slip activity for the 50–50 nm HCP Mg/BCC Nb composites (Taken with permission from Ardeljan et al. 2018)

In the Nb phase, the model indicates that {112} slip dominates the slip activity over {110} slip in both loading directions. Since the {112} and {110} slip modes had the same CLS parameters ($\tau_0^s, A^s, f^s$), the higher propensity for {112} slip is a consequence of texture.

Slip activity calculations in the Mg and Nb also provide insight into the origin of plastic anisotropy, namely, the higher composite flow stress seen in the normal direction than the transverse direction (Fig. 17c). Generally, Nb is the stronger material compared to Mg. The model finds that Nb accommodates proportionally more applied strain in the normal loading than parallel loading, which would alone cause normal loading to give rise to a higher flow stress than parallel loading. In addition, more ⟨c + a⟩ slip in the Mg phase is activated under normal loading than under parallel loading. In Mg, pyramidal ⟨c + a⟩ slip has the highest CRSS compared to prismatic and basal slip, and thus, its higher activity would explain the larger yield in normal loading than parallel loading.

Last, an interesting finding is the source of strain hardening. Since the texture did not change, geometric hardening would not have contributed. Analysis of slip activity and the deformation of the layers indicate that the strain hardening is due to a combination of a reduction in $h$ and change in texture. The hardening is higher in the normal case than parallel case, since in the normal case, the reduction in $h$ during deformation is more severe.

# 4    Part III: Areas for Mesoscale Processing-Microstructure Modeling

Thus far, this chapter has discussed the latest developments in 3D, mesoscale CP-based modeling techniques, and some recent advancements in the ability to model the role mesoscale microstructural features, such as grain (orientation, size, and shape), grain boundaries, twin boundaries, and interface properties, play in the generation of local stresses states and deformation mechanisms during mechanical processing or straining. As is common in modeling efforts, models and their capability to inform and predict could benefit from a number of extensions.

For the mesoscale techniques reviewed here, useful future modifications concern the ability to model the effects of microstructure on the onset (when), formation (where), and migration of boundaries within an explicit 3D microstructure. The list of possible stress-induced phenomena that would introduce internal microstructural boundaries during processing is long: nucleation and growth of voids, development of localized slip bands and shear bands, phase transformations, recrystallization, grain growth, and twins. Also included in this area is the need to account for interactions between dislocations and twins with grain boundaries, interfaces, or precipitates. The number of possible defect/interface reactions that could affect microstructural evolution is long as well. While many of these aspects have been studied intensively and widely using atomic-scale simulation, representing the role of such highly resolved atomic-scale reactions into a mesoscale model is lacking. At the same time, new strategies are needed to overcome the numerical issues involved in dynamically creating and evolving boundaries in explicit microstructure mesoscale models. Below we briefly describe some new and upcoming strategies for tackling these areas.

The heterogeneous stress and microstructural states in a strained material can lead to the dissipation of energy in the form of voids. The study of void growth in crystals and crystalline boundaries has been an intensely studied topic in atomic-scale simulation, discrete dislocation dynamics, phase-field, finite elements, and continuum mechanics (Needleman 1972; Koplik and Needleman 1988; Gurson 1977; Tvergaard and Needleman 1984). Yet, among the modeling techniques available to date, using explicit microstructural CP-based modeling in the study of voids in 3D single crystals and polycrystals is relatively new. In single crystals, these techniques are beginning to probe the effects of void size and orientation on the strain to initiate void growth and on growth rates (Wan et al. 2005; Yu et al. 2010; Ha and Kim 2010; Potirniche et al. 2006; Liu et al. 2007; Han et al. 2013b;

Yerra et al. 2010; Schacht et al. 2003; Mbiakop et al. 2015; Srivastava and Needleman 2015; Pushkareva et al. 2016). In polycrystals, the effect of anisotropic plasticity on void growth rates (or onset of collapse) in tension (or compression) loading states can be accessed (Lebensohn and Cazacu 2012). Looking forward, incorporation of voids would lead to the ability to design processes that do not prematurely lead to failure or process-aware microstructures that postpone fracture.

Deformation twinning is another stress-induced deformation mechanism that introduces boundaries inside crystals. The various schemes for modeling twin reorientation reviewed here, however, were designed for mean-field techniques, and therefore, none of them are intended to explicitly account for twin boundary formation. Up to now, advancing CP models for discrete twins with a 3D microstructure has been hindered by the lack of 3D microstructural representation codes. Very recently, a few microstructural models of discrete twin domains within individual grains have been developed (Ardeljan et al. 2015b, 2017; Abdolvand and Wilkinson 2016; Cheng and Ghosh 2017). With these techniques, studies on the local stress fields and dislocation activity around twin lamellae have shed light on the effects of size and local grain neighborhoods on the propensity for twin growth Fig. 20. Although the twin lamellae can be explicitly inserted, expanding the twin domain dynamically in CPFE can be computationally intensive. CP method advancement for growing discrete twins with a 3D microstructure has been hindered by lack of efficient geometry and mesh manipulation tools and the fact that the mechanisms underlying the nucleation and propagation of twin domains and the relevant driving forces are still being studied.

Many high-performance alloys (e.g., Mg alloys, Al alloys, and superalloys) contain multiple phases of different sizes, which introduce many heterophase
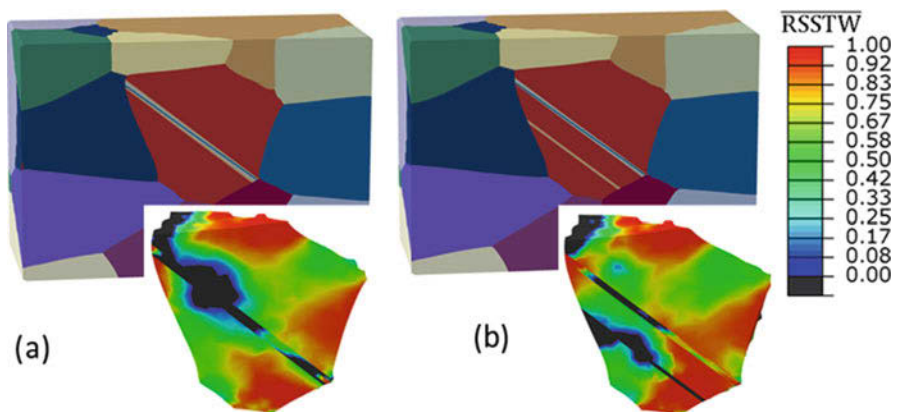


**Fig. 20** Twin formation and thickening with respect to the surrounding neighboring grains in the 3D model of α-U under compression in the z-direction with (**a**) one 5% twin lamella and (**b**) one of 2% and another of 3% volume fraction. The insert shows the distributions of normalized resolved shear stresses along the most dominant twin plane in the twin direction for the red interior grain (Taken with permission from Ardeljan et al. 2015b)

boundaries. For example, Inconels contain γ′, γ″, and δ precipitates, varying in chemical composition, size, and shape. Interactions of discrete dislocations with precipitates have been studied via atomistic, discrete dislocation dynamics, and phase-field methods (Beyerlein and Hunter 2016; Gao et al. 2015). With CP-based models, the effect of precipitates is commonly represented via a change in the CRSS (Shenoy et al. 2008; Ghorbanpour et al. 2017; Nugmanov et al. 2018; Lentz et al. 2015), which models phenomenologically the effect of precipitate size on whether the dislocation can shear it or not. CP can also provide a way to calculate the mechanical stress, strain, and strain rate fields within the individual crystals as a result of elastic and slip on crystallographic slip systems in each phase as well as the differences in deformation across the boundary. Yet to benefit the design of advanced alloys by microstructure control, advancements in CP modeling toward incorporating the role of dislocation/interface interactions in slip activity would be recommended.

## 5    Conclusions

Over the years, mesoscale materials modeling approaches have been built to elucidate the role played by microstructure, e.g., grain shape, grain orientation, grain size, internal grain boundaries, and interfaces, in processing-microstructure-property relationships of polycrystalline materials. Many high-performance polycrystalline materials of intense interest for future engineering applications are inherently or microstructurally complex, warranting advancements in such mesoscale deformation models beyond the standard versions. This chapter aims to highlight the latest developments in 3D, mesoscale CP-based modeling techniques, and the challenges they have overcome. In particular, some recent advancements in the ability to model the role that mesoscale microstructural features, such as grain (orientation, size, and shape), grain boundary, and interface properties, play in the development of local stress states and deformation mechanisms during mechanical processing or straining are described and discussed. Examples cover studies that have employed and extended mesoscale 3D, full-field spatial resolved computational models to investigate processing-microstructure relationships in boundary-dominant materials. These include the influence of bimetal interfaces on microstructural evolution. Also presented and discussed are advancements made to mesoscale techniques to study microstructure-property relationships in boundary-dominant materials. Particular problems addressed include the effects of grain size and shape and layer and twin thickness on dislocation emission and propagation. The chapter ends with a discussion of possible directions and strategies future mesoscale modeling efforts could take to advance capability and value in the design of advanced materials.

# References

Abdolvand H, Wilkinson AJ (2016) Assessment of residual stress fields at deformation twin tips and the surrounding environments. Acta Mater 105:219–231

Al-Fadhalah K et al (2005) Modeling texture evolution during rolling of a Cu–Nb multilayered system. Philos Mag 85(13):1419–1440

Anderoglu O et al (2008) Thermal stability of sputtered Cu films with nanoscale growth twins. J Appl Phys 103(9):094322

Ardeljan M et al (2015a) A study of microstructure-driven strain localizations in two-phase polycrystalline HCP/BCC composites using a multi-scale model. Int J Plast 74:35–57

Ardeljan M et al (2015b) Explicit incorporation of deformation twins into crystal plasticity finite element models. Comput Methods Appl Mech Eng 295:396–413

Ardeljan M, Beyerlein IJ, Knezevic M (2017) Effect of dislocation density-twin interactions on twin growth in AZ31 as revealed by explicit crystal plasticity finite element modeling. Int J Plast 99:81–101

Ardeljan M, Knezevic M, Jain M, Pathak S (2018) Room temperature deformation mechanisms of Mg/Nb nanolayered composites. J Mater Res 33:1311

Barrett CS, Massalski MA (1966) Structure of metals. McGraw-Hill, New York

Beyerlein IJ, Hunter A (2016) Understanding dislocation mechanics at the meso-scale using phase field dislocation dynamics. Philos Trans R Soc A 374:20150166

Beyerlein IJ, Mayeur JR (2015) Mesoscale investigations for the evolution of interfaces in plasticity. Curr Opin Solid State Mater Sci 19(4):203–211

Beyerlein IJ, Tomé CN (2008) A dislocation-based constitutive law for pure Zr including temperature effects. Int J Plast 24(5):867–895

Beyerlein IJ, McCabe RJ, Tomé CN (2011) Effect of microstructure on the nucleation of deformation twins in polycrystalline high-purity magnesium: a multi-scale modeling study. J Mech Phys Solids 59(5):988–1003

Beyerlein IJ et al (2012) Structure–property–functionality of bimetal interfaces. JOM 64(10):1192–1207

Beyerlein IJ et al (2013a) Radiation damage tolerant nanomaterials. Mater Today 16(11):443–449

Beyerlein IJ, Wang J, Zhang R (2013b) Interface-dependent nucleation in nanostructured layered composites. APL Mater 1(3):032112

Beyerlein IJ, Wang J, Zhang R (2013c) Mapping dislocation nucleation behavior from bimetal interfaces. Acta Mater 61(19):7488–7499

Beyerlein IJ et al (2013d) Interface-driven microstructure development and ultra high strength of bulk nanostructured Cu-Nb multilayers fabricated by severe plastic deformation. J Mater Res 28(13):1799–1812

Beyerlein IJ, Zhang X, Misra A (2014) Growth twins and deformation twins in metals. Annu Rev Mater Res 44(1):329–363

Beyerlein I, Demkowicz MJ, Misra A, Uberuaga P (2015) Defect-interface interactions. Prog Mater Sci 74:125–210

Bronkhorst CA, Kalidindi SR, Anand L (1991) An experimental and analytical study of the evolution of crystallographic texturing in FCC materials. Texture Microstruct 14:1031–1036

Bronkhorst CA, Kalidindi SR, Anand L (1992) Polycrystalline plasticity and the evolution of crystallographic texture in FCC metals. Philos Trans R Soc Lond Ser A Math Phys Eng Sci 341(1662):443–477

Carpenter JS, McCabe RJ, Mayeur JR, Mara NA, Beyerlein IJ (2014) Interface-driven plasticity: the presence of an interface affected zone in metallic lamellar composites. Adv Eng Mat 17:109–114. https://doi.org/10.1002/adem.201400210

Carpenter JS et al (2015a) Bulk texture evolution of nanolamellar Zr–Nb composites processed via accumulative roll bonding. Acta Mater 92:97–108

Carpenter JS et al (2015b) The suppression of instabilities via biphase interfaces during bulk fabrication of nanograined Zr. Mater Res Lett 3(1):50–57

Champion Y et al (2001) Fabrication of bulk nanostructured materials from metallic nanopowders: structure and mechanical behaviour. Scr Mater 44(8–9):1609–1613

Chen J, Lu L, Lu K (2006) Hardness and strain rate sensitivity of nanocrystalline Cu. Scr Mater 54(11):1913–1918

Chen XH, Lu L, Lu K (2007) Electrical resistivity of ultrafine-grained copper with nanoscale growth twins. J Appl Phys 102(8):083708

Cheng J, Ghosh S (2017) Crystal plasticity finite element modeling of discrete twin evolution in polycrystalline magnesium. J Mech Phys Solids 99:512–538

Christian JW, Mahajan S (1995) Deformation twinning. Prog Mater Sci 39(1–2):1–157

Dalla Torre F, Van Swygenhoven H, Victoria M (2002) Nanocrystalline electrodeposited Ni: microstructure and tensile properties. Acta Mater 50(15):3957–3970

Dao M, Lu L, Shen YF, Suresha S (2006) Strength, strain-rate sensitivity and ductility of copper with nanoscale twins. Acta Mater 54(20):5421–5432

Das D, Samanta A, Chattopadhyay PP (2006) Mechanical properties of bulk ultrafine-grained copper. Synth Reactivity Inorg Metal Org Nano Metal Chem 36(2):221–225

Delannay L, Jacques PJ, Kalidindi SR (2006) Finite element modeling of crystal plasticity with grains shaped as truncated octahedrons. Int J Plast 22(10):1879–1898

DREAM.3D Version 4.2 (2013) BlueQuartz Software. Springboro

Ebrahimi F et al (1999) Mechanical properties of nanocrystalline nickel produced by electrodeposition. Nanostruct Mater 11(3):343–350

El-Sherik AM, Erb U, Palumbo G, Aust KT (1992) Deviations from hall-petch behaviour in as-prepared nanocrystalline nickel. Scr Metall Mater 27(9):1185–1188

Foreman AJE (1967) The bowing of a dislocation segment. Philos Mag 15(137):1011–1021

Gao S, Fivel M, Ma A, Hartmaier A (2015) Influence of misfit stresses on dislocation glide in single crystal superalloys: a three-dimensional discrete dislocation dynamics study. J Mech Phys Solids 76:276–290

Ghorbanpour S et al (2017) A crystal plasticity model incorporating the effects of precipitates in superalloys: application to tensile, compressive, and cyclic deformation of Inconel 718. Int J Plast 99(Supplement C):162–185

Godon A et al (2010) Effects of grain orientation on the Hall–Petch relationship in electrodeposited nickel with nanocrystalline grains. Scr Mater 62(6):403–406

Groeber MA, Jackson MA (2014) DREAM. 3D: a digital representation environment for the analysis of microstructure in 3D. Integr Mater Manuf Innov 3(1):5

Gurson AL (1977) Continuum theory of ductile rupture by void nucleation and growth: part I—yield criteria and flow rules for porous ductile media. J Eng Mater Technol 99(1):2–15

Ha S, Kim K (2010) Void growth and coalescence in f.c.c. single crystals. Int J Mech Sci 52(7):863–873

Han W et al (2013a) Design of radiation tolerant materials via interface engineering. Adv Mater 25(48):6975–6979

Han X et al (2013b) A yield function for single crystals containing voids. Int J Solids Struct 50(14):2115–2131

Hughes GD et al (1986) Hall-Petch strengthening for the microhardness of twelve nanometer grain diameter electrodeposited nickel. Scr Metall 20(1):93–97

Hunter A, Leu B, Beyerlein IJ (2018) Mesoscale modeling of slip transmission through biphase interfaces. J Mater Sci 53:5584–5603

Jérusalem A, Dao M, Suresh S, Radovitzky R (2008) Three-dimensional model of strength and ductility of polycrystalline copper containing nanoscale twins. Acta Mater 56(17):4647–4657

Kalidindi SR (1998) Incorporation of deformation twinning in crystal plasticity models. J Mech Phys Solids 46(2):267–271

Kang K, Wang J, Beyerlein IJ (2012) Atomic structure variations of mechanically stable fcc-bcc interfaces. J Appl Phys 111(5):053531

Khan A, Farrokh B, Takacs L (2008) Compressive properties of Cu with different grain sizes: sub-micron to nanometer realm. J Mater Sci 43(9):3305–3313

Kikuchi S et al (1997) Mechanical properties of Ag-Ni super-laminates produced by rolling. Mater Sci Eng A 234–236:1114–1117

Knezevic M et al (2014) Three dimensional predictions of grain scale plasticity and grain boundaries using crystal plasticity finite element models. Comput Methods Appl Mech Eng 277:239–259

Kocks UF, Tomé CN, Wenk H-R (1998) Texture and anisotropy. Cambridge University Press, Cambridge, UK

Koplik J, Needleman A (1988) Void growth and coalescence in porous plastic solids. Int J Solids Struct 24(8):835–853

Kumar KS, Van Swygenhoven H, Suresh S (2003a) Mechanical behavior of nanocrystalline metals and alloys1. Acta Mater 51(19):5743–5774

Kumar KS et al (2003b) Deformation of electrodeposited nanocrystalline nickel. Acta Mater 51(2):387–405

Kumar A et al (2017) An atomic-scale modeling and experimental study of ⟨c+a⟩ dislocations in Mg. Mater Sci Eng A 695(Supplement C):270–278

Lebensohn RA, Cazacu O (2012) Effect of single-crystal plastic deformation mechanisms on the dilatational plastic response of porous polycrystals. Int J Solids Struct 49(26): 3838–3852

Lebensohn RA, Tomé CN (1993) A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys. Acta Metall Mater 41(9):2611–2624

Lebensohn RA, Tomé CN (1994) A self-consistent viscoplastic model: prediction of rolling textures of anisotropic polycrystals. Mater Sci Eng A 175(1–2):71–82

Lebensohn RA, Liu Y, Ponte Castañeda P (2004) On the accuracy of the self-consistent approximation for polycrystals: comparison with full-field numerical simulations. Acta Mater 52(18):5347–5361

Lebensohn RA, Tomé CN, Castaneda PP (2007) Self-consistent modelling of the mechanical behaviour of viscoplastic polycrystals incorporating intragranular field fluctuations. Philos Mag 87(28):4287–4322

Lebensohn RA, Rollett AD, Suquet P (2011) Fast Fourier transform-based modeling for the determination of micromechanical fields in polycrystals. JOM 63(3):13–18

Lentz M et al (2015) Effect of age hardening on the deformation behavior of an Mg–Y–Nd alloy: in-situ X-ray diffraction and crystal plasticity modeling. Mater Sci Eng A 628:396–409

Li L et al (2009) The stress–strain response of nanocrystalline metals: a quantized crystal plasticity approach. Acta Mater 57(3):812–822

Li N et al (2012) Direct observations of confined layer slip in Cu/Nb multilayers. Microsc Microanal 18(5):1155–1162

Lieberman EJ et al (2016) Microstructural effects on damage evolution in shocked copper polycrystals. Acta Mater 116:270–280

Liu WH et al (2007) Simulation of void growth and coalescence behavior with 3D crystal plasticity theory. Comput Mater Sci 40(1):130–139

Liu B et al (2010) Comparison of finite element and fast Fourier transform crystal plasticity solvers for texture prediction. Model Simul Mater Sci Eng 18(8):085005

Lu L et al (2004) Ultrahigh strength and high electrical conductivity in copper. Science 304(5669):422–426

Lu L et al (2009a) Revealing the maximum strength in nanotwinned copper. Science 323(5914):607–610

Lu L et al (2009b) Size dependence of rate-controlling deformation mechanisms in nanotwinned copper. Scr Mater 60(12):1062–1066

Mara N, Beyerlein I (2014) Review: effect of bimetal interface structure on the mechanical behavior of Cu–Nb fcc–bcc nanolayered composites. J Mater Sci 49(19):6497–6516

Mara NA et al (2008) Deformability of ultrahigh strength 5 nm Cu/Nb nanolayered composites. Appl Phys Lett 92(23):231901

Mayeur J et al (2014) The influence of grain interactions on the plastic stability of heterophase interfaces. Materials 7(1):302–322

Mayeur JR et al (2015) Incorporating interface affected zones into crystal plasticity. Int J Plast 65:206–225

Mbiakop A, Constantinescu A, Danas K (2015) An analytical model for porous single crystals with ellipsoidal voids. J Mech Phys Solids 84:436–467

Meyers MA, Mishra A, Benson DJ (2006) Mechanical properties of nanocrystalline materials. Prog Mater Sci 51(4):427–556

Mika DP, Dawson PR (1998) Effects of grain interaction on deformation in polycrystals. Mater Sci Eng A 257(1):62–76

Mirkhani H, Joshi SP (2011) Crystal plasticity of nanotwinned microstructures: a discrete twin approach for copper. Acta Mater 59(14):5603–5617

Mirkhani H, Joshi SP (2014) Mechanism-based crystal plasticity modeling of twin boundary migration in nanotwinned face-centered-cubic metals. J Mech Phys Solids 68:107–133

Misra A, Hoagland R, Kung H (2004) Thermal stability of self-supported nanolayered Cu/Nb films. Philos Mag 84(10):1021–1028

Misra A, Hirth J, Hoagland R (2005) Length-scale-dependent deformation mechanisms in incoherent metallic multilayered composites. Acta Mater 53(18):4817–4824

Misra A et al (2007) The radiation damage tolerance of ultra-high strength nanolayered composites. JOM 59(9):62–65

Monclús MA et al (2013) Optimum high temperature strength of two-dimensional nanocomposites. APL Mater 1(5):052103

Needleman A (1972) Void growth in an elastic-plastic medium. J Appl Mech 39(4):964–970

Nizolek T et al (2014) Processing and deformation behavior of bulk Cu–Nb nanolaminates. Metallogr Microstruct Anal 3(6):470–476

Nizolek T et al (2015) Enhanced plasticity via kinking in cubic metallic nanolaminates. Adv Eng Mater 17(6):781–785

Nizolek T et al (2016) Tensile behavior and flow stress anisotropy of accumulative roll bonded Cu-Nb nanolaminates. Appl Phys Lett 108(5):051903

Nugmanov D et al (2018) Origin of plastic anisotropy in (ultra)-fine-grained Mg–Zn–Zr alloy processed by isothermal multi-step forging and rolling: experiments and modeling. Mater Sci Eng A 713:81–93

Pan QS, Lu L (2014) Strain-controlled cyclic stability and properties of Cu with highly oriented nanoscale twins. Acta Mater 81:248–257

Pan QS, Lu QH, Lu L (2013) Fatigue behavior of columnar-grained Cu with preferentially oriented nanoscale twins. Acta Mater 61(4):1383–1393

Partridge PG (1967) The crystallography and deformation modes of hexagonal close-packed metals. Metall Revised 12:169

Pathak S et al (2017) Strong, ductile, and thermally stable bcc-Mg nanolaminates. Sci Rep 7(1):8264

Potirniche GP et al (2006) Lattice orientation effects on void growth and coalescence in fcc single crystals. Int J Plast 22(5):921–942

Proust G et al (2009) Modeling the effect of twinning and detwinning during strain-path changes of magnesium alloy AZ31. Int J Plast 25(5):861–880

Pushkareva M et al (2016) Three-dimensional investigation of grain orientation effects on void growth in commercially pure titanium. Mater Sci Eng A 671:221–232

Ritz H, Dawson P (2008) Sensitivity to grain discretization of the simulated crystal stress distributions in FCC polycrystals. Model Simul Mater Sci Eng 17(1):015001

Sarma GB, Dawson PR (1996) Effects of interactions among crystals on the inhomogeneous deformations of polycrystals. Acta Mater 44(5):1937–1953

Schacht T, Untermann N, Steck E (2003) The influence of crystallographic orientation on the deformation behaviour of single crystals containing microvoids. Int J Plast 19(10):1605–1626

Schuh CA, Nieh TG, Yamasaki T (2002) Hall–Petch breakdown manifested in abrasive wear resistance of nanocrystalline nickel. Scr Mater 46(10):735–740

Schwaiger R et al (2003) Some critical experiments on the strain-rate sensitivity of nanocrystalline nickel. Acta Mater 51(17):5159–5172

Shen TD et al (1995) On the elastic moduli of nanocrystalline Fe, Cu, Ni, and Cu–Ni alloys prepared by mechanical milling/alloying. J Mater Res 10(11):2892–2896

Shen YF et al (2005) Tensile properties of copper with nano-scale twins. Scr Mater 52(10):989–994

Shenoy M, Tjiptowidjojo Y, McDowell D (2008) Microstructure-sensitive modeling of polycrystalline IN 100. Int J Plast 24(10):1694–1730

Srivastava A, Needleman A (2015) Effect of crystal orientation on porosity evolution in a creeping single crystal. Mech Mater 90:10–29

Taylor GI (1938) Plastic strain in metals. J Inst Met 62:307–324

Tvergaard V, Needleman A (1984) Analysis of the cup-cone fracture in a round tensile bar. Acta Metall 32(1):157–169

Van Houtte P (1978) Simulation of the rolling and shear texture of brass by the Taylor theory adapted for mechanical twinning. Acta Metall Mater 26(4):591–604

Wan J, Yue Z, Lu Z (2005) Casting microporosity growth in single-crystal superalloys by a three-dimensional unit cell analysis. Model Simul Mater Sci Eng 13(6):875

Wang J, Hoagland RG, Hirth JP, Misra A (2008) Atomistic simulations of the shear strength and sliding mechanisms of copper–niobium interfaces. Acta Mater 56(13):3109–3119

Wang J et al (2012) Slip transmission across fcc/bcc interfaces with varying interface shear strengths. Acta Mater 60(4):1503–1513

Wang J et al (2013) Characterizing interface dislocations by atomically informed Frank-Bilby theory. J Mater Res 28(13):1646–1657

Xiao C et al (2001) Tensile behavior and fracture in nickel and carbon doped nanocrystalline nickel. Mater Sci Eng A 301(1):35–43

Xie C et al (2015) The impact of twin lamella thickness distribution on strength and endurance limit in nanotwinned copper. Mech Mater 84:91–99

Yasuna K et al (2000) Formation of nanoscale Fe/Ag multilayer by repeated press-rolling and its layer thickness dependence of magnetoresistance. Mater Sci Eng A 285(1–2):412–417

Yerra S et al (2010) Void growth and coalescence in single crystals. Int J Solids Struct 47(7):1016–1029

Yoo MH (1981) Slip, twinning, and fracture in hexagonal close-packed metals. Metall Mater Trans A 12:409–418

Yoshinaga H, Obara T, Morozumi S (1973) Twinning deformation in magnesium compressed along the C-axis. Mater Sci Eng 12(5–6):255–264

You Z et al (2013) Plastic anisotropy and associated deformation mechanisms in nanotwinned metals. Acta Mater 61(1):217–227

You ZS, Lu L, Lu K (2011) Tensile behavior of columnar grained Cu with preferentially oriented nanoscale twins. Acta Mater 59(18):6927–6937

Youngdahl CJ et al (1997) Compressive yield strengths of nanocrystalline Cu and Pd. Scr Mater 37(6):809–813

Youssef KM et al (2004) Ultratough nanocrystalline copper with a narrow grain size distribution. Appl Phys Lett 85(6):929–931

Yu Q, Hou N, Yue Z (2010) Finite element analysis of void growth behavior in nickel-based single crystal superalloys. Comput Mater Sci 48(3):597–608

Yuan R, Beyerlein IJ, Zhou C (2015) Emergence of grain-size effects in nanocrystalline metals from statistical activation of discrete dislocation sources. Acta Mater 90:169–181

Yuan R, Beyerlein IJ, Zhou C (2016a) Coupled crystal orientation-size effects on the strength of nano crystals. Sci Rep 6:26254

Yuan R, Beyerlein IJ, Zhou C (2016b) Statistical dislocation activation from grain boundaries and its role in the plastic anisotropy of nanotwinned copper. Acta Mater 110(Supplement C): 8–18

Zecevic M et al (2016) Origin of texture development in orthorhombic uranium. Mater Sci Eng A 665:108–124

Zhang X et al (2005) Thermal stability of sputter-deposited 330 austenitic stainless-steel thin films with nanoscale growth twins. Appl Phys Lett 87(23):233116

Zhang Y et al (2007) High strength and high electrical conductivity in bulk nanograined Cu embedded with nanoscale twins. Appl Phys Lett 91(21):211901

Zheng S et al (2013) High-strength and thermally stable bulk nanolayered composites due to twin-induced interfaces. Nat Commun 4:1696

Zhu YT, Liao XZ, Wu XL (2008) Deformation twinning in bulk nanocrystalline metals: experimental observations. JOM 60(9):60

Zhu L et al (2011) Modeling grain size dependent optimal twin spacing for achieving ultimate high strength and related high ductility in nanotwinned metals. Acta Mater 59(14):5544–5557

Zhu L et al (2015) Analysis of the twin spacing and grain size effects on mechanical properties in hierarchically nanotwinned face-centered cubic metals based on a mechanism-based plasticity model. J Mech Phys Solids 76:162–179

# Adaptive Physics Refinement at the Microstructure Scale

# 52

Timothy C. Germann

## Contents

**Abstract**

A long-sought goal of computational materials science and engineering has been a simulation framework that spans all necessary length and time scales, potentially from electronic structure to structural engineering, providing the appropriate level of physics fidelity where needed and enabling the user to trade off accuracy and computational time in an optimal manner. Analogous to adaptive mesh refinement methods that dynamically (and automatically) coarsen and refine a computational mesh based on local requirements, *adaptive physics refinement* methods utilize higher-fidelity physics models as needed, e.g., replacing a phenomenological constitutive model with a direct polycrystal

T. C. Germann (✉)
Theoretical Division (T-1), Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: tcg@lanl.gov

plasticity simulation. While there have been several demonstrations of similar concurrent multiscale methods over the past 20–30 years, only now is a more general capability becoming viable due to advances in algorithms and computer architectures and middleware. In this chapter, we briefly review this history, focusing on two methods in particular: the heterogeneous multiscale method and adaptive sampling. The computational workflow, data, and runtime requirements of these methods are used to identify key enabling technologies that have recently gained widespread adoption, including task-based programming models, heterogeneous computer architectures, database, and machine learning algorithms.

# 1 Introduction

Multiscale modeling techniques, in which different levels of theory are combined to provide higher accuracy in regions where required, while utilizing more computationally efficient models elsewhere, have a long history in computational science. Perhaps the most notable of these is the hybrid quantum mechanics/molecular mechanics (QM/MM) method for computational chemistry (Warshel and Levitt 1976), whose developers were awarded the 2013 Nobel Prize in Chemistry for "the development of multiscale models for complex chemical systems." In the 1990s, similar techniques were developed and applied to solid-state materials by coupling atomistic and finite element methods. A common prototypical application was to fracture mechanics, combining a long-range nonlinear stress field described by a continuum finite element model with an atomistic description of a crack tip (Kohlhoff et al. 1991) or even extended to a third level of theory by using a tight-binding model for bond rupture in the macroscopic, atomistic, ab initio dynamics (MAAD) technique (Abraham et al. 1998; Broughton et al. 1999). For the reader interested in these and related techniques, including the quasicontinuum (Tadmor et al. 1996; Miller and Tadmor 2002) and coarse-grained molecular dynamics (Rudd and Broughton 1998) methods, a number of comprehensive reviews exist (Lu and Kaxiras 2005; Bernstein et al. 2009; Miller and Tadmor 2009).

# 2 The Heterogeneous Multiscale Method

Motivated by these and other multiscale, multiphysics approaches, E and Engquist introduced the broader heterogeneous multiscale method (HMM) framework, in analogy to the "homogeneous" multigrid methods in which the same physical model is employed at different scales (Weinan and Engquist 2003a, b; Weinan et al. 2003, 2007). In most previous multiscale methods, different models are applied to different physical regions, for instance, a molecular dynamics (MD) region and a finite element (FE) region, with some overlap, handshaking, or bridging region where consistency is maintained. In contrast, HMM uses detailed microscopic models to provide the necessary closure relations for general macroscopic conservation laws. As with the representative atoms ("repatoms") of the quasicontinuum

method, fine-scale (typically MD) models are employed for each element where a macroscopic constitutive model may either be unknown, too expensive to evaluate, or locally invalid, e.g., due to crystal defects or chemical reactions. In a typical approach, a finite element or finite volume model enforces continuum conservation laws for mass, momentum, and energy, while local MD simulations for each element, at each time step, provide average momentum and energy fluxes as statistical time averages, assuming a separation of scales and ergodic dynamical behavior.

## 3    Scale-Bridging Strategies Using Polycrystal Plasticity Models and Adaptive Sampling

Plastic anisotropy (i.e., a mechanical response which depends upon the orientation of the constituent grains) is an important property that is closely tied to a polycrystalline material's microstructure and determines how that material deforms under extreme mechanical conditions. Traditionally, macroscale finite element simulations represent this effect using (static) phenomenological plasticity models which have been parameterized from experiments and/or a more realistic polycrystal plasticity model, for a fixed microstructure. For scenarios involving the coupling of plastic anisotropy with microstructure evolution, direct multiscale embedding has been employed (e.g., Segurado et al. 2012), analogous to the original HMM approach (Weinan et al. 2007) in which every material element resorts to a lower length-scale calculation, every time step. However, the computational cost of such a direct embedding becomes prohibitive as one moves from relatively simple homogenization assumptions to models with higher physical fidelity and correspondingly higher computational expense.

Microstructural effects can be explicitly considered in micromechanical models, with different degrees of fidelity and numerical efficiency, using polycrystal plasticity formulations. The simplest and most efficient Taylor model assumes identical deformation throughout all single crystal grains in a representative volume element (RVE) of the material, typically consisting of a number of grains ranging from 50 (providing a coarse representation) to 1000 (refined), and it is insensitive to the spatial arrangement of the grains within the RVE. Given the strain of each grain (which is known from the applied boundary conditions), the Taylor model gives the stress response as an average of the stresses at the grain level, obtained by solving one system of five nonlinear equations for each grain. A typical refined Taylor computation therefore involves the solution of 1000 decoupled $5 \times 5$ systems of nonlinear equations, requiring on the order of milliseconds on a single modern CPU core.

This Taylor scheme, while computationally efficient, unrealistically neglects interactions between grains. To address this limitation, more elaborate homogenization schemes like the viscoplastic self-consistent (VPSC) model (Lebensohn and Tomé 1993) have been proposed. VPSC considers every grain as an ellipsoidal inclusion embedded in a homogenized effective medium (HEM). Like the Taylor

model, VPSC results do not depend on the spatial arrangement of grains. The HEM properties and the strains and stresses of the grains are not known a priori, but must be obtained iteratively, involving a weighted average of the responses at the grain level. A typical refined VPSC calculation involves 3–5 iterations, each requiring the solution of 1000 coupled $10 \times 10$ systems of nonlinear equations, i.e., less than a second on a single core.

Both the Taylor and VPSC models have an important limitation in common: they are mean-field approaches in which the morphology of the grains is either neglected (as in Taylor) or a simplified grain shape is assumed (ellipsoidal, as in VPSC); and the influence of the specific neighborhood of a given grain is either neglected (as in Taylor) or averaged out (by means of the HEM assumption, as in VPSC). Alternatively, a full-field viscoplastic model based on Fast Fourier Transforms (VPFFT) (Lebensohn 2001) is able to explicitly treat complex grain morphologies and topologies, in a relatively efficient way compared with other full-field approaches such as the crystal plasticity finite element method (CPFEM). The VPFFT approach is based on the use of FFTs to efficiently calculate convolution integrals between a Green's function associated with a reference homogeneous medium and a suitably defined polarization field. Since the polarization depends precisely on the unknown micromechanical fields, an iterative scheme is used. All fields are mapped onto a 3-D regular grid, typically consisting of a number of voxels ranging from $16 \times 16 \times 16$ (for a coarse representation) to $64 \times 64 \times 64$ (refined) or potentially even greater. A typical refined VPFFT calculation involves 10–20 iterations, each iteration consisting of direct and inverse FFTs calculated over the regular grid, and the solution of one $5 \times 5$ system of nonlinear equations per voxel, requiring several minutes on a single processor core. This computation time can be reduced to a few seconds at the expense of losing fidelity by using a coarse grid (see Table 1).

To overcome the high computational cost of directly embedding such polycrystal plasticity models, Barton et al. (2008) developed a scale-bridging methodology that employs adaptive sampling (AS), mitigating the computational expense of direct embedding by building a database of material response as a function of polycrystal microstructure as a simulation proceeds. Whenever a response is needed (typically for each material element, each time step), this database is queried, and if existing database entries are sufficiently close to the new state, the result is interpolated using kriging or another interpolation method. (How one defines "sufficiently close"

**Table 1** Single-core CPU time measured for different polycrystal plasticity models and fidelities of the microstructure description

|  | Mean field representation | |
| --- | --- | --- |
|  | Coarse (50 grains) | Refined (1000 grains) |
| Taylor | 1.8 ms | 35 ms |
| VPSC | 5 ms | 100 ms |
|  | Full field representation | |
|  | Coarse ($16^3$ grid) | Refined ($64^3$ grid) |
| VPFFT | 2.3 s | 148 s |

requires uncertainty quantification analyses, as discussed below in Sect. 4.3.) The original AS implementation was based on the relatively simple Taylor model, which does not account for texture evolution by slip and twinning. Recently, Barton et al. (2015) unified the VPSC model with a discrete harmonic (DH) representation of crystallographic texture and the AS embedding strategy, integrating all three models in the ALE3D finite element code (Noble et al. 2017). The DH-based representation of texture and its evolution was the key to define a database that is representative of the microstructure, is of manageable size, and can account for the rapid texture evolution produced by deformation twinning. This implementation achieved two or more orders of magnitude in wall-clock speedup, compared with direct interrogation of VPSC by the finite element model.

In addition to the increased computational expense of the higher-fidelity poly-crystal plasticity models, two additional computational considerations should be noted. First, the increased model complexity (and reality) is associated with an increase in the number of input parameters required to describe the material microstructure. For instance, the recent ALE3D-AS-DH-VPSC implementation (Barton et al. 2015) just mentioned represents the material texture with a DH expansion; adding higher-order harmonics provides an increasingly accurate description but also introduces many more coefficients. The application of this approach to predict texture-induced anisotropy effects on the final shape of heavily textured Ti bars bent along different directions required a 60-dimensional input space for a sixth-order expansion. The dimensionality is even greater for the spatially resolved VPFFT model, so one must consider the trade-off between accuracy and dimensionality for reduced dimensionality approximations.

This increased dimensionality is in part the original motivation and promise of AS methods: precomputing and parameterizing a model that covers this space are computationally prohibitive and unnecessary since, in practice, the points queried during any particular calculation tend to fall on a low-dimensional manifold of this many-dimensional parameter space (Barton et al. 2011). However, the input space dimensionality introduces additional burdens on both the database lookup and the interpolation method. In addition to these challenges for the AS layer, the history dependence also requires that the coarse-scale model (e.g., ALE3D) tracks the evolving microstructure at each material point.

On future exascale (and beyond) supercomputers, one may be able to afford a spatially resolved VPFFT polycrystal plasticity model, but such calculations are today prohibitive due to computational cost (several minutes per fine-scale model, whereas a coarse-scale time step should ideally take no more than a few seconds) and dimensionality challenges. As an intermediate step between the simple Taylor model and the still-intractable VPFFT model, current state-of-the-art efforts use the intermediate mean-field VPSC plasticity model. The single-core runtime (10–100 ms) is roughly comparable to what we expect each VPFFT calculation to take on an exascale-era node, and the intermediate input space dimensionality (tens of parameters) provides a smooth migration pathway from the simple Taylor model (with only a few parameters) to VPFFT (with hundreds to thousands of input parameters).

# 4    Recent Extensions and Applications

## 4.1    Spatial Adaptive Sampling

The key insight that enables adaptive sampling to offer substantial performance gains is that during any given simulation, the fine-scale microstructure queries follow a low-dimensional manifold of a high-dimensional input space, gradually varying in space and time, aside from shocks and other discontinuities. This enables a frequent reuse and interpolation of recently computed and accessed fine-scale evaluations and kriging models. However, as just mentioned, the potentially high-dimensional input space introduces challenges both in quickly accessing previously computed nearby results and in developing reliable interpolation models. Instead of interpolating over this input parameter space, another option is to interpolate over *physical* space, which of course is at most three dimensional. This approach has been proposed and successfully demonstrated in Rouet-Leduc et al. (2014) and is illustrated in Fig. 1. In these initial demonstrations, previously computed fine-scale results are not stored, nor are kriging interpolation models developed; instead, at each macroscale time step, the required constitutive data is reconstructed by interpolating an automatically determined subset of fine-scale simulations on the $d \leq 3$-dimensional subdomain. As time progresses, one can predict which fine-scale sample points will be required, for a given target accuracy. These authors demonstrated that spatial adaptive sampling can reduce the number of fine-scale simulations (and thus the total computational cost) by three orders of magnitude, compared to direct HMM.

## 4.2    Practical Applications to Molecular Solids

Following the move of two of the key developers of the original adaptive sampling methodology (Arsenlis et al. 2006; Barton et al. 2008, 2011; Knap et al. 2008)
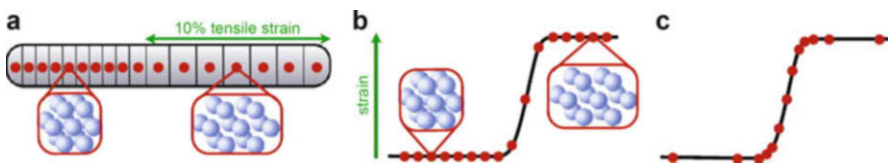


**Fig. 1** Schematic illustration of the spatial adaptive sampling scheme applied to the simple problem of elastodynamic shock propagation in one dimension. (**a**) Shock propagation dynamics occurs on macroscopic space and time scales. Red dots denote representative microscopic regions of the system, each of which is modeled by $\simeq 1000$ atoms of a defect-free fcc Cu crystal. (**b**) The heterogeneous multiscale method (HMM) integrates conservation laws to determine the dynamical evolution of the macroscopic fields. Constitutive data is provided by the stochastic estimates of microscopic molecular dynamics simulations, performed on a regular grid. (**c**) In spatial adaptive sampling, the location of microscopic simulations is dynamically adapted to increase the accuracy and efficiency of the HMM simulation (Rouet-Leduc et al. 2014). Permission from Elsevier

from Lawrence Livermore National Laboratory to the Army Research Laboratory, it has been adopted and further developed there in recent years (Knap et al. 2015, 2016; Barnes et al. 2017a, b). Their primary application has been to model low-velocity Taylor anvil impact tests for molecular solids, in particular energetic materials below the threshold for chemical reactions (Barnes et al. 2017b). Their concurrent hierarchical multiscale (HMS) scale-bridging framework combines three loosely coupled components: a microscopic model, a macroscopic model, and the scale-bridging "glue." At the continuum scale, ALE3D (Noble et al. 2017) is an arbitrary Lagrangian-Eulerian hydrodynamics code. At the fine scale, an interface to the LAMMPS code (Plimpton 1995) has been developed that provides on-demand calculation of material response given an arbitrary state point (e.g., density and internal energy). ALE3D and this LAMMPS integrated materials engine (LIME) (Barnes et al. 2017a) are then coupled via a stand-alone evaluation module that is responsible for collecting microscopic model evaluation requests from the microscopic model, scheduling their evaluation given available computational resources, and communicating their results back to the macroscopic model.

## 4.3   Algorithmic Challenges

Adaptive sampling methods introduce several algorithmic challenges, including how one identifies "nearby" previous fine-scale evaluations that can be used for interpolation, how this interpolation from the microstructure-response database is done, and how one determines whether interpolation is sufficiently accurate, or if a new fine- scale model evaluation must be launched.

The problem of finding neighboring points in a possibly high-dimensional space has been intensely addressed by the computer vision community (Shakhnarovich et al. 2006). A key decision is whether one requires an exact nearest neighbor search or whether an approximate nearest neighbor search suffices. The original adaptive sampling implementations stored a database of kriging models rather than the more numerous set of fine-scale evaluation results. Consequently, an exact neighbor search (using a metric-tree database) was used to find the (single) kriging model whose centroid was closest to the queried point. This approach has the advantage of enabling the reuse of kriging models that are very expensive to construct but also introduces complications including (1) how to use new fine-scale model evaluations to update a kriging model; (2) when to merge neighboring kriging models or, more typically, split an overextended kriging model into two separate models; and (3) how to smoothly switch from one kriging model to another for intermediate query points. As a result of these complications, other difficulties due to dimensionality and scalability, and estimates of the database size for the two approaches, more recent efforts have evaluated an alternative implementation in which the fine-scale model outputs are stored in the database directly, rather than the subsequent interpolation models.

Consequently, an approximate nearest neighbor search will suffice to identify enough neighboring points to compute an interpolated estimate. Among approxi-

mate nearest neighbor search algorithms, locality sensitive hashing (LSH) (Andoni and Indyk 2008) is a leading method. Despite offering good (theoretical) scaling with dimension and data size, LSH requires many hyper-parameters to tune and is not suited for the inhomogeneous data expected for our microstructure-response data sets. Instead, one promising alternative approach would use a forest of search tress, specifically either k-d trees, as implemented in the Fast Library for Approximate Nearest Neighbors (FLANN) (Muja and Lowe 2014), or a forest of v-p trees, as implemented in the Proximity Forest (O'Hara and Draper 2013). Preliminary tests by our group show both methods to work well, as measured by the fraction of exact nearest neighbors correctly identified and by the number of trees required.

Kriging, or Gaussian process regression, models were originally designed for geospatial interpolation and used in prior adaptive sampling work. However, kriging scales poorly with dimensionality and will be unlikely to work as we move from the few-dimensional input parameter space of the simpler proof-of-concept models (Barton et al. 2008, 2011; Knap et al. 2008) toward the $O(10^3 - 10^4)$-dimensional space of other applications. As a result, simpler local linear or quadratic interpolants, which will work well for smooth functions such as the microstructure-response surface, have been evaluated. Error estimates, which are provided (with some assumptions) by kriging algorithms, can be obtained by sampling different sets of neighbors, e.g., from differently randomized search trees.

This error estimate can then be used to decide whether the interpolation result is sufficiently accurate to use or a new fine-scale evaluation is required. A naïve (and common) approach is to reject the interpolated result if this error estimate exceeds a chosen threshold. Applying such a uniform threshold at each local microstructure-response query is well defined but not necessarily what is optimal from a global perspective. For instance, as a material is deforming, the microstructure evolution and shape of the material may be more (or less) sensitive to errors in the underlying plasticity description in some regions at any given time, and thus a tighter (or looser) error tolerance would produce the same overall response. In this global perspective, one needs to define a metric (or metrics) to evaluate the accuracy of a simulation, as by comparing the final material shapes resulting from a dynamic loading experiment.

This leads to the well-defined goal of increasing the accuracy of a predicted quantity of interest (QoI). When a computational model is used to predict a material property or process, the QoI can be thought of as a functional of the constitutive laws being used. In a general case, the QoI depends on a set of input parameters (e.g., geometry, environmental conditions, etc.) and a set of constitutive functions (e.g., stress-strain relationship). Optimal UQ (OUQ) strives to find the optimal bounds on quantities of interest given the information at hand and to enable the optimal design of computational experiments (Owhadi et al. 2013). To address model uncertainty, Strachan et al. (2013) have recently developed a functional uncertainty quantification (UQ) approach that enables the estimation of uncertainties in the predictions due to the use of low-fidelity materials models. One practical use of functional UQ has been for MD simulations performed using the Lennard-Jones

potential, inferring thermodynamic properties corresponding to different potentials without requiring additional simulations (Reeve and Strachan 2017).

# 5   Computational Requirements for Task-Based Scale-Bridging

From a computer/computational science perspective, the exploration of dynamic, asynchronous, concurrent approaches is critical as we move toward exascale, since the massive concurrency involved will introduce processor performance variability, operating system (OS) jitter, and a plethora of soft, hard, and silent errors that will make tightly coupled bulk synchronous approaches increasingly unviable. Instead of this traditional parallel programming approach where the programmer is responsible for assigning equal amounts of work to each MPI process, more loosely coupled asynchronous models where the work is overdecomposed, and the runtime system responsible for distributing and balancing the work given task and data dependencies, will enable a more efficient use of resources, as well as much improved fault tolerance.

Based upon these concerns about the physics fidelity and architectural trends toward massive concurrency, heterogeneity, and asynchronous, dynamic approaches rather than traditional bulk synchronous parallel models, adaptive physics refinement such as HMM or AS, whereby a coarse-scale simulation dynamically spawns fine-scale simulations as needed, are attractive. This strategy is crucial for capturing how the macroscale, bulk response is influenced by microstructural detail. Thus, in a high strain-rate loading problem, a finite element calculation may spawn finer-scale crystal plasticity or atomistic models as needed when the available empirical constitutive model is inadequate. The basic workflow of such a concurrent scale-bridging application workflow is illustrated in Fig. 2 and involves a coarse-scale simulation, dynamically spawned fine-scale simulation tasks, a database for storing the results of fine-scale tasks, and an adaptive sampling layer which queries the database, interpolates results, and decides when to spawn new fine-scale tasks. This approach to increased physics fidelity via embedded scale-bridging imposes several requirements upon the underlying computational software stack, including task-based programming and execution models, runtimes, and databases.

This approach has generated an increasing amount of interest in the material mechanics community, where several groups are actively developing various related methods, alternatively referred to as kriging, surrogate model, or metamodel construction. In particular, computational materials programs such as Advanced Simulation and Computing/Physics and Engineering Models (ASC/PEM), the Materials Genome Initiative (MGI), Integrated Computational Materials Engineering (ICME), the Advanced Manufacturing Initiative, etc. all recognize that improving physical fidelity, rather than extending time or length scales by brute force, is the greatest driver for increased computational power. In addition, such concurrent scale-bridging approaches are being developed in several other materials science contexts, not to mention other computational science domains.
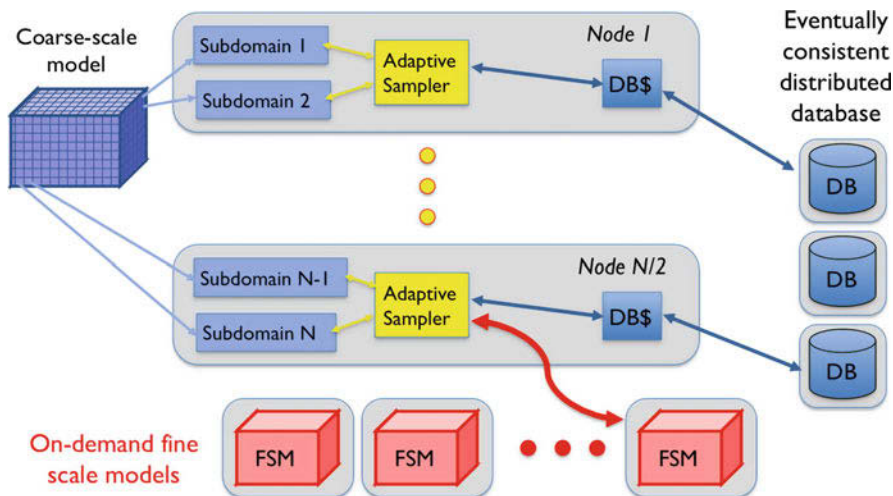
**Fig. 2** Components of a scale-bridging application. A coarse-scale model runs across multiple nodes, each utilizing adaptive sampling to either spawn fine-scale models (FSM) as needed or interpolate previous fine-scale responses stored in a database (DB), consisting of an on-node cache (DB$) and a distributed database

Examples include interpolation-based MD potentials (Bartók et al. 2010; Trott et al. 2014), including a recent demonstration with on-the-fly database construction and interpolation (Li et al. 2015) and on-the-fly kinetic Monte Carlo approaches (Henkelman and Jónsson 2001; Trushin et al. 2005; Xu et al. 2011). As other researchers have noted (Barton et al. 2011; Knap et al. 2015, 2016; Alowayyed et al. 2017), such methods have computational workflows that are drastically different from traditional, tightly coupled bulk synchronous programming (BSP) approaches to scientific computing, as we will detail below.

## 5.1    Workflow Requirements

The structure and workflow of one recent prototype application, the task-based scale-bridging code (Tabasco) which was demonstrated by the Exascale Co-design Center for Materials in Extreme Environments (ExMatEx) on the heterogeneous Trinity supercomputer, is shown in Fig. 3. Out of the five individual components (A-E) comprising the application, two (A and E) represent the two physics scales: the LULESH coarse-scale (A) and VPSC fine-scale (E) codes. (LULESH Karlin et al. (2013) is a proxy application for ALE3D, representing the basic Lagrangian hydrodynamics computation without the full complexity of a production code.) The remaining components (B-D) implement the heart of the adaptive sampling approach and are what dramatically reduce the number of expensive fine-scale calls needed during each time step.
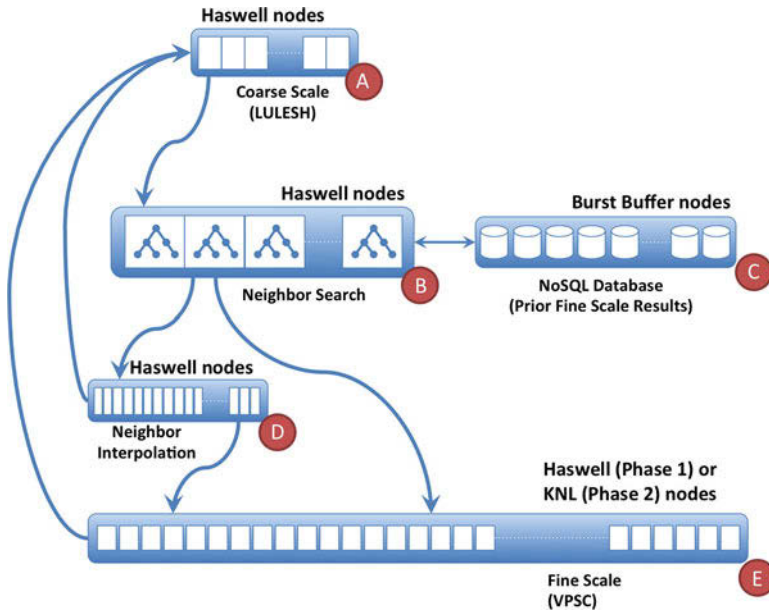
**Fig. 3** Structure and workflow of the task-based scale-bridging code (Tabasco) prototype as implemented on the heterogeneous Trinity supercomputer

These components each require a different number of nodes, and this node usage can vary with time. Additionally, as shown in the diagram, each component is best suited to a particular type of node (on Trinity, either Haswell, KNL, or burst buffer). In addition, some components are long-lived and stateful (A–C) while others are ephemeral and stateless – running for a short time only, returning their results to another component, and exiting. This dynamic and asynchronous computing approach requires a task-based programming model. We have explored a number of such models as candidates to implement our prototype. These programming models are responsible for allocating resources to components, scheduling and managing the execution of tasks within components, and communicating results between components.

In order to better understand the application flow, we will walk through the execution of a single time step (again, refer to Fig. 3).

1. At the beginning of the time step, each mesh element of the LULESH coarse-scale computation (A) requires a material stress to advance its state. This stress (and other quantities, such as texture evolution parameters) must be computed by the VPSC fine-scale component. In a brute force approach, each coarse-scale cell would spawn a fine-scale call (E) to provide the required stress for that cell.
2. Adaptive sampling improves dramatically on the brute force approach by dynamically launching a fine-scale call only for mesh elements where it is required. Due

to the evolving nature of the physics, many previously computed fine-scale call values continue to be useful into the future on mesh cells at a similar state in the computation.

3. Three components (B-D) are used to implement the adaptive sampling method. As each fine-scale request from LULESH (A) arrives, an approximate neighbor search (B) is executed. This search, distributed and executing on many nodes concurrently, searches a set of trees for neighbors of the query point. As the search traverses the trees, it retrieves previously computed fine-scale results for these neighboring points from a NoSQL – or key/value – database (C).

4. If a sufficient number of neighbor points are found, a fast interpolation (D) can be performed in an attempt to generate a suitable approximation to the required stress (and texture evolution). If the interpolation fails, or if there are an insufficient number of neighbor points, a new VPSC fine-scale call (E) is dynamically spawned.

5. If interpolation is successful, the stress is returned to the requesting LULESH coarse-scale (A) element. If a new VPSC fine-scale call is required, its value is similarly returned to its requesting element, and in addition it is queued for eventual insertion into the search tree structure and database, where it can be used for subsequent interpolation attempts. (Note that, while our adaptive sampling approach reduces overall fine-scale calls by 90%, the remaining 10% will still consume the lion's share of the prototype's execution time – especially in the case of more complex response models like VPSC. Acceleration of these fine-scale calls, using KNL nodes on Trinity or GPUs on Sierra, provides a performance leveraging opportunity to substantially reduce this remaining execution time.)

6. When all requesting coarse-scale cell requests have been fulfilled, new points queued for insertion are added to the search trees and database, coarse-scale mesh element updates are executed, any needed coarse-scale neighbor exchanges are performed, and the computation advances to the next time step.

The use of adaptive sampling as a scale-bridging technique presents significant challenges in the management of computing resources, including the scheduling of tasks performing fine-scale model evaluations and the management of the dynamic, distributed database containing previous model evaluations. Because the corresponding workflows are difficult to predict a priori, the ability to generate example workflows is the first step in the characterization of more general scenarios.

## 5.2    Runtime Requirements

Many, if not most, of today's scientific simulation applications are developed using a fairly limited set of software technologies: a standard programming language such as C, C++, or FORTRAN (possibly along with node-level acceleration APIs like CUDA or OpenMP and solver libraries such as Trilinos), MPI for communication, and a static scheduler, such as Slurm or Moab, to execute the computation on the machine. Should a developer need to load balance their computation, they need to

provide this functionality themselves. Similarly, fault tolerance requires programmers to periodically write data to disk (or use in-memory checkpointing Glosli et al. 2007) for later recovery. Likewise, communication patterns are generally fixed – dynamic communication patterns must be designed and implemented by the programmer.

Modern runtime systems can provide much of this functionality as a collection of independent "system services." These services support "programming in the large" – coupling multiple diverse components of a dynamic multiscale computation and orchestrating its execution on the system. These services, by category of functionality, are:

- Scheduling: concurrent, asynchronous, adaptively executing computational components, launched on-the-fly, and exiting when complete.
- Discovery: locating system resources based on application-supplied requirements (e.g., provide a list of all nodes with GPU accelerators).
- Messaging: setup and teardown, on-the-fly, dynamic, adaptive communication links between components of the calculation.
- Caching: services for temporarily storing data, perhaps in-memory, and retrieving it from anywhere in the computation. Caching can help prevent duplicate computation or store data to be used for recovery from faults. Caching can also be used for communication. Instead of sending messages, processes can store their data in the cache for retrieval by other processes.
- Fault detection: working with the application, operating system, and hardware, detects faults in the system and provides facilities for application notification or automated restart.

In general, there are two ways in which these services (or subsets of them) are implemented. First, there are distributed monolithic systems that are closely tied to a programming language or model (e.g., Charm++, X10, Chapel, CnC, or Erlang). These systems include a runtime component that implements features of the programming models such as scheduling, communication, data distribution, etc. (These system-level runtimes should not be confused with low-level runtimes provided by component-level programming languages such as CUDA or OpenMP, which often provide the same conceptual features but at a much finer level of granularity.) Second, there is a more loosely coupled approach that uses various single-function software, usually open source, to build an integrated, dynamic system. This approach is closer to what industrial developers (e.g., Netflix, Facebook, LinkedIn, Google, etc.) use to build cloud- or web-based scalable systems.

To explore this design space, the Exascale Co-Design Center for Materials in Extreme Environments developed a simplified CoHMM "proxy application" for the basic HMM method (ExMatEx 2015a) and implemented it with a variety of programming models, runtimes, and database backends (Roehm et al. 2015; Pavel et al. 2015). As a monolithic approach, the Erlang programming language provides the desired features in a single language and runtime system. For the fine-scale molecular dynamics tasks, the simplified CoMD proxy was used (ExMatEx

2015b). An alternative, cloud-based approach used a more diverse set of tools including Apache ZooKeeper (for discovery, scheduling, and process tracking), node.js to manage the overall execution of the code, and multiple NoSQL databases (MongoDB, memcached, Riak, Couchbase) to cache data for fault tolerance. In both cases, the implementations simply managed the dynamic coupling between the coarse- and fine-scale components of CoHMM. All of the mathematical computation was done in the component-level APIs that those components used (e.g., MPI or OpenMP).

The 2013 Los Alamos IS&T Co-Design Summer School, comprised of six students with backgrounds in both computer science as well as physics, refined the CoHMM and CoMD proxy apps and used them to test a variety of runtime systems from industry and academia. They evaluated software that acts as the "glue" between the coarse- and fine-scale components of CoHMM, whose simplicity allowed us to investigate a wide swath of software technologies – both monolithic and cloud based. These evaluations focused on some of the primitive features described above: scheduling, communication, caching, and fault tolerance. Various schedulers were tested against how well they supported dynamic and adaptive task scheduling.

In addition, fault tolerance was implemented using two approaches. In the first, process-level case, the runtime system detects the crash of a fine-scale CoMD process, and the system restarts it from its initial conditions (potentially on another node) without crashing the entire application. In the second case, an in-memory database (again, Redis) is used to periodically cache particle positions from each CoMD process. If the runtime system detects a crash, the CoMD process is restarted from the conditions encapsulated in the most recently cached particle positions – not from the initial conditions. These developments also enhanced our original test applications; fault tolerance was added to the Erlang version, and our cloud version now uses the proven Redis database.

## 5.3    Database Requirements

Achieving maximum performance for Tabasco relies on a database of fine-scale calculations to enable multidimensional interpolation (kriging) and avoid additional, costly, fine-scale computations. Initial explorations by Barton, Knap, and colleagues at LLNL (Barton et al. 2008, 2011; Knap et al. 2008) used a hand-coded database to cache fine-scale computations. Since then, distributed NoSQL databases (Sadalage and Fowler 2012) have become ubiquitous in web and cloud applications and are attractive (primarily the key-value store variations) based on their ease of use, community support, and potential performance.

As with the exploration of task-based programming models and runtime systems, CoHMM has been used to explore APIs and functionality of key-value stores. CoHMM is a tractable proxy for these explorations and matches the algorithmic structure of the Tabasco target scale-bridging application. Keys are formed by concatenating truncated floating point parameters from the coarse-scale code, effectively binning the fine-scale values. Alternatively, distributed tree-based

approximate nearest neighbor (ANN) techniques are in principle better suited to adaptive sampling access patterns and also remain amenable to distributed database implementations.

Multiple versions of CoHMM have been implemented, primarily using the Redis (2018) NoSQL database. Redis was chosen for its ease of installation, simple interface, and adequate performance. This integration of Redis into CoHMM successfully achieved reductions on the order of 90% in the number of fine-scale calculations. A number of issues require further exploration before committing to any particular database: performance, non-transactional semantics, and support on target petascale platforms.

The performance of a single-node Redis database is more than adequate for CoHMM, where fine-scale calls take several seconds to execute, and thus the database performance is not stressed. As we move to much faster fine-scale calculations (on the order of 10 ms), absolute performance will become a challenge. However, early experience indicates that many of these potential databases will provide adequate performance for this problem. For example, a single-node Redis database on a MacBook Pro can support about 70,000 queries per second, or about 14 $\mu$s per query. Other distributed databases (e.g., Aerospike 2017 and RAMCloud Ousterhout et al. 2010) can support queries at rates exceeding 500,000 per second. This throughput, though, is measured at the database server. Ultimate throughput must account for network latency which may, over TCP/IP, be the primary bottle-neck. RAMCloud's use of Infiniband network technology could support very low latencies if required but may be more difficult to use than commodity databases. Finally, we note that actual production applications may actually employ more accurate and costly fine-scale models, thereby reducing latency requirements on the database. By "skeletonizing" CoHMM, replacing real coarse-scale (e.g., LULESH) and fine-scale (e.g., CoMD, VPSC, or VPFFT) codes with tunable null-value functions, one may experiment more broadly with database performance by dialing up or down the costs of these fine-scale tasks.

A second potential issue with NoSQL databases is that they generally don't support transactional semantics but are instead "eventually consistent." This means that the database only guarantees that a write will eventually be readable by a client. It is possible that a query for an in-flight write will fail. Compared to traditional, synchronized databases, NoSQL databases are easier to implement, faster, and more scalable because of these semantics. Eventual consistency doesn't appear to be a problem for our scale-bridging application. If we suffer a miss on a fine-scale query, we can simply rerun the fine-scale computation for that query. In practice, we haven't noticed problems within CoHMM due to eventual consistency. However, without a sophisticated in-band logging capability and post execution analysis, it is difficult to calculate a miss rate. To date, no NoSQL databases provide this logging framework.

A third and potentially serious issue is that of system support for NoSQL databases. These databases were designed with commercial datacenters in mind (i.e., web and cloud applications). They typically use TCP/IP for networking and are installed into operating environments that are tolerant of diverse, and continuously

evolving, system software. Current HPC systems typically use Infiniband for networking and enforce a strict limitation on available system software, which may heavily restrict the set of usable NoSQL databases and how they are used. The ExMatEx project explored this issue in two ways. First, they successfully installed Redis and ran CoHMM on multiple HPC platforms including LANL's Darwin cluster (a heterogeneous prototyping platform), LLNL's Vulcan (IBM BlueGene/Q), and LLNL's Cab (Intel Xeon-based). Second, they explored possible mitigation approaches in case the database cannot be run in a distributed fashion (e.g., due to lack of TCP/IP). In this case, node-local copies of the database may be run with periodic MPI-based, cross-system merges. Another mitigation approach may be to use an LRU eviction policy on local databases to keep their memory requirements tractable. Such experience with a nontraditional scientific computing use of databases in an HPC environment may influence long-term requirements for future extreme scale systems.

In addition to reducing required computation, NoSQL databases can be used for a number of other optimizations or additional functionality. While most NoSQL databases run in an ephemeral, in-memory mode – they disappear at the conclusion of a calculation – some are able to persist to NVRAM or rotational storage (as in the use of a burst-buffer backed database on Trinity). In this case, once fine-scale response database has been built and captured, it can then be used subsequently on similar problems (e.g., design parameter studies for additive manufacturing). New fine-scale calls would only be required when the current parameter study deviates significantly from the one (or many) that preceded it and pre-populated the microstructure-response database, which is mostly used in a query-only mode. Second, the captured database could be used for fault tolerance. Instead of checkpointing the problem at each time step, a persistent database could be used to advance a restart without ever executing a single fine-scale call. The ExMatEx project experimented with this possibility early in the project by caching the particle positions of a long-running CoMD task, thereby enabling a restart upon detection of a failure. Third, in situ analytics can be performed along with the computation by concurrently accessing the database using the same task-based and system capabilities as the computation does.

## 6    Future Directions

As this chapter has hopefully conveyed, there is a promising future for adaptive physics refinement methods in materials modeling, as well as several practical technical challenges, spanning materials science, applied mathematics, and computer science. Among these, modern machine learning techniques offer one attractive alternative for the traditional kriging interpolation models, which as mentioned suffer as the input parameter space dimensionality increases to hundreds, thousands, or beyond. The increasing heterogeneity of computer architectures, both at the processor and system scales, and asynchronous task-based programming models

and runtimes, is both a challenge to conventional scientific software codes that have been developed in the massively parallel era of the past 20–30 years and a tremendous opportunity for novel approaches to scientific simulation as we enter the exascale computing era.

# References

Abraham FF, Broughton JQ, Bernstein N, Kaxiras E (1998) Spanning the length scales in dynamic simulation. Comput Phys 12(6):538–546

Aerospike, Inc (2017) Aerospike NoSQL database. http://aerospike.com

Alowayyed S, Groen D, Coveney PV, Hoekstra AG (2017) Multiscale computing in the exascale era. J Comput Sci 22:15–25

Andoni A, Indyk P (2008) Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. Commun ACM 51(1):117–122

Arsenlis A, Barton N, Becker R, Rudd R (2006) Generalized in situ adaptive tabulation for constitutive model evaluation in plasticity. Comput Methods Appl Mech Eng 196(1–3):1–13

Barnes BC, Leiter KW, Becker R, Knap J, Brennan JK (2017a) LAMMPS integrated materials engine (LIME) for efficient automation of particle-based simulations: application to equation of state generation. Model Simul Mater Sci Eng 25(5):055006

Barnes BC, Spear CE, Leiter KW, Becker R, Knap J, Lisal M, Brennan JK (2017b) Hierarchical multiscale framework for materials modeling: equation of state implementation and application to a Taylor Anvil impact test of RDX. In: Chau R, Germann T, Oleynik I, Peiris S, Ravelo R, Sewell T (eds) Shock compression of condensed matter – 2015, Tampa. AIP conference proceedings, vol 1793

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403

Barton NR, Knap J, Arsenlis A, Becker R, Hornung RD, Jefferson DR (2008) Embedded polycrystal plasticity and adaptive sampling. Int J Plast 24(2):242–266

Barton NR, Bernier JV, Knap J, Sunwoo AJ, Cerreta EK, Turner TJ (2011) A call to arms for task parallelism in multi-scale materials modeling. Int J Numer Methods Eng 86:744–764

Barton NR, Bernier JV, Lebensohn RA, Boyce DE (2015) The use of discrete harmonics in direct multi-scale embedding of polycrystal plasticity. Comput Methods Appl Mech Eng 283:224

Bernstein N, Kermode JR, Csányi G (2009) Hybrid atomistic simulation methods for materials systems. Rep Prog Phys 72(2):026501

Broughton JQ, Abraham FF, Bernstein N, Kaxiras E (1999) Concurrent coupling of length scales: methodology and application. Phys Rev B Condens Matter Mater Phys 60(4):2391–2403

ExMatEx (2015a) CoHMM: a co-design proxy application for the Heterogeneous Multiscale Method (HMM). https://github.com/exmatex/CoHMM

ExMatEx (2015b) CoMD: classical molecular dynamics proxy application. https://github.com/ECP-copa/CoMD

Glosli JN, Richards DF, Caspersen KJ, Rudd RE, Gunnels JA, Streitz FH (2007) Extending stability beyond CPU millennium: a micron-scale atomistic simulation of Kelvin-Helmholtz instability. In: SC '07: Proceedings of the 2007 ACM/IEEE conference on supercomputing. ACM, New York, pp 1–11

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115:9657

Karlin I, Bhatele A, Keasler J, Chamberlain BL, Cohen J, DeVito Z, Haque R, Laney D, Luke E, Wang F, Richards D, Schulz M, Still C (2013) Exploring traditional and emerging parallel programming models using a proxy application. In: 27th IEEE international parallel & distributed processing symposium (IEEE IPDPS 2013), Boston

Knap J, Barton NR, Hornung RD, Arsenlis A, Becker R, Jefferson DR (2008) Adaptive sampling in hierarchical simulation. Int J Numer Methods Eng 76(4):572–600

Knap J, Spear CE, Borodin O, Leiter KW (2015) Advancing a distributed multi-scale computing framework for large-scale high-throughput discovery in materials science. Nanotechnology 26(43):434004

Knap J, Spear C, Leiter K, Becker R, Powell D (2016) A computational framework for scale-bridging in multi-scale simulations. Int J Numer Methods Eng 108(13):1649–1666

Kohlhoff S, Gumbsch P, Fischmeister HF (1991) Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model. Philos Mag A 64(4):851–878

Lebensohn RA (2001) N-site modeling of a 3D viscoplastic polycrystal using fast Fourier Transform. Acta Mater 49(14):2723–2737

Lebensohn RA, Tomé C (1993) A self-consistent approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys. Acta Metall Mater 41:2611

Li Z, Kermode JR, De Vita A (2015) Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. Phys Rev Lett 114:096405

Lu G, Kaxiras E (2005) Overview of multiscale simulations of materials. In: Rieth M, Schommers W (eds) Handbook of theoretical and computational technology, vol X. American Scientific Publishers, Stevenson Ranch, pp 1–33

Miller R, Tadmor E (2002) The quasicontinuum method: overview, applications and current directions. J Compu-Aided Mater Des 9(3):203–239

Miller RE, Tadmor EB (2009) A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Model Simul Mater Sci Eng 17(5):053001

Muja M, Lowe DG (2014) Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans Pattern Anal Mach Intell 36(11):2227–2240

Noble CR, Anderson AT, Barton NR, Bramwell JA, Capps A, Chang MH, Chou JJ, Dawson DM, Diana ER, Dunn TA, Faux DR, Fisher AC, Greene PT, Heinz I, Kanarska Y, Khairallah SA, Liu BT, Margraf JD, Nichols AL, Nourgaliev RN, Puso MA, Reus JF, Robinson PB, Shestakov AI, Solberg JM, Taller D, Tsuji PH, White CA, White JL (2017) ALE3D: an arbitrary Lagrangian-Eulerian multi-physics code. https://doi.org/10.2172/1361589

O'Hara S, Draper BA (2013) Are you using the right approximate nearest neighbor algorithm? In: 2013 IEEE workshop on applications of computer vision (WACV), Clearwater Beach, pp 9–14

Ousterhout J, Agrawal P, Erickson D, Kozyrakis C, Leverich J, Mazières D, Mitra S, Narayanan A, Parulkar G, Rosenblum M, et al (2010) The case for RAMClouds: scalable high-performance storage entirely in DRAM. ACM SIGOPS Oper Syst Rev 43(4):92–105

Owhadi H, Scovel C, Sullivan TJ, McKerns M, Ortiz M (2013) Optimal uncertainty quantification. SIAM Rev 55(2):271–345

Pavel RS, McPherson AL, Germann TC, Junghans C (2015) Database assisted distribution to improve fault tolerance for multiphysics applications. In: Proceedings of the 2nd international workshop on hardware-software co-design for high performance computing, Co-HPC '15. ACM, New York, pp 4:1–4:8

Plimpton S (1995) LAMMPS: large-scale atomic/molecular massively parallel simulator. http://lammps.sandia.gov

Redislabs (2018) Redis database. http://redis.io

Reeve ST, Strachan A (2017) Error correction in multi-fidelity molecular dynamics simulations using functional uncertainty quantification. J Comput Phys 334:207–220

Roehm D, Pavel R, Barros K, Rouet-Leduc B, McPherson A, Germann T, Junghans C (2015) Distributed database kriging for adaptive sampling ($D^2KAS$). Comput Phys Commun 192: 138–147

Rouet-Leduc B, Barros K, Cieren E, Elango V, Junghans C, Lookman T, Mohd-Yusof J, Pavel RS, Rivera AY, Roehm D, McPherson AL, Germann TC (2014) Spatial adaptive sampling in multiscale simulation. Comput Phys Commun 185(7):1857–1864

Rudd RE, Broughton JQ (1998) Coarse-grained molecular dynamics and the atomic limit of finite elements. Phys Rev B 58:R5893–R5896

Sadalage PJ, Fowler M (2012) NoSQL distilled: a brief guide to the emerging world of polyglot persistence, 1st edn. Addison-Wesley Professional, Boston

Segurado J, Lebensohn RA, LLorca J, Tomé CN (2012) Multiscale modeling of plasticity based on embedding the viscoplastic self-consistent formulation in implicit finite elements. Int J Plast 28:124

Shakhnarovich G, Darrell T, Indyk P (2006) Nearest-neighbor methods in learning and vision: theory and practice (Neural Information Processing). The MIT Press, Cambridge

Strachan A, Mahadevan S, Hombal V, Sun L (2013) Functional derivatives for uncertainty quantification and error estimation and reduction via optimal high-fidelity simulations. Model Simul Mater Sci Eng 21(6):065009

Tadmor EB, Ortiz M, Phillips R (1996) Quasicontinuum analysis of defects in solids. Philos Mag A 73(6):1529–1563

Trott CR, Hammond SD, Thompson AP (2014) SNAP: strong scaling high fidelity molecular dynamics simulations on leadership-class computing platforms. In: Proceedings of the 29th international conference on supercomputing, ISC 2014, vol 8488. Springer, New York, pp 19–34

Trushin O, Karim A, Kara A, Rahman TS (2005) Self-learning kinetic Monte Carlo method: application to Cu(111). Phys Rev B 72:115401

Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 103(2):227–249

Weinan E, Engquist B (2003a) The heterogeneous multiscale methods. Commun Math Sci 1(1): 87–132

Weinan E, Engquist B (2003b) Multiscale modeling and computation. Not Am Math Soc 50(9):1062–1070

Weinan E, Engquist B, Huang Z (2003) Heterogeneous multiscale method: a general methodology for multiscale modeling. Phys Rev B 67(9):092101

Weinan E, Engquist B, Li X, Ren W, Vanden-Eijnden E (2007) Heterogeneous multiscale methods: a review. Commun Comput Phys 2(3):367–450

Xu H, Osetsky YN, Stoller RE (2011) Simulating complex atomistic processes: on-the-fly kinetic Monte Carlo scheme with selective active volumes. Phys Rev B 84:132103

# 53

# Synchrotron Capabilities to Understand Microstructure of Additively Manufactured Materials: Challenges and Opportunities for Modeling and Simulations

Anthony D. Rollett

## Contents

**Abstract**

From the perspective of modeling and simulation, additive manufacturing is an unambiguously multiscale problem. Regardless of whether the 3D printing is accomplished via melting, or polymerization, or with binders, the scale of the process is submillimeter, which means that dozens to thousands of layers are accumulated while making a part. Variations in geometry mean that the path followed by the light or electron beam (except in the case of whole layer-based illumination) result in highly variable time intervals between successive overlapping heat inputs. Particularly in the case of processes that melt powders, this can result in deviations from the expected heat input that lead to defects. Taking microstructure to be the totality of the structure of crystal(s) and defects,

A. D. Rollett (✉)
Department of Materials Science and Engineering, Carnegie Mellon University,
Pittsburgh, PA, USA
e-mail: rollett@cmu.edu

this means that using simulation to predict microstructure requires calculations at multiple scales: it is not feasible to simulate microstructure development at the submillimeter scale with, e.g., grains, orientations, and pores, when the part extends to centimeters in all dimensions. Synchrotron radiation is well suited to probing the unit processes involved in additive manufacturing, and so a focus on the submillimeter scale of materials processing provides a useful framework for evaluating needs and prospects for modeling and simulation.

# 1    Introduction

This chapter is intended to introduce the reader to the techniques that are most useful for characterizing additively manufactured materials and the implications of recent results for modeling and simulation. The techniques emphasize crystalline materials, which mean mainly metals and ceramics. Similar challenges exist for additively manufactured polymers, and bioprinting is particularly interesting because of the rapid development that existed at the time of writing. Although synchrotron radiation has seen less development for these applications, it is expected to be equally impactful in the future. The techniques that are discussed include x-ray micro-computed tomography (xμCT), wide-angle X-ray scattering (WAXS), high-energy diffraction microscopy (HEDM), residual stress (RS), and dynamic X-ray radiography (DXR). The various investigations have brought, e.g., new knowledge about unexpected precipitation behaviors in Ni-based alloys and keyhole formation in selective laser melting systems.

Additive manufacturing (AM) covers a substantial range of technologies for making prototypes, intermediate forms, and final parts in nearly all materials. Several books and review articles are available that, to varying degrees, explain how the technologies work (Gibson et al. 2010). In a simplified view, the AM field has been moving from mainly prototyping objects with polymers to manufacturing final parts in both polymers and metals that require only minimal finishing (Wohlers 2017). Complexity of the design and part count reduction, along with rapid introduction to market, all favor AM methods. Focusing on metal parts, small size is helpful because the build rate in the dominant powder bed machines is of the order of 3 mm/h, which means that tall builds have to run for a few days. Increasing the build rate is, of course, a major thrust of the machine makers. Other technologies such as wire feed, which is based on either electron beam or laser welding, allow much larger parts albeit with less resolution and more significant post-processing. Binder jet technologies avoid melting with its inevitable residual stress but require binder removal and sintering. Other technologies are being developed such as precision metal droplet deposition (Murr and Johnson 2017).

Simulation of such processes is self-evidently multiscale even for just the deposition phase: metal powder bed, for example, requires melting many meters of melt lines, each of which is of order $200 \times 100\,\mu m$ in cross section (width × depth). Adjacency of the melt lines allied with the melting through each layer to (at least) the layer beneath means that the thermal history of each location requires several lines

and layers to be simulated. Simulating the entire build sequence is also necessary, however, for computing residual stress, distortion, as well as heat buildup that affects the actual thermal history required for predicting microstructure. Accordingly, it is common to consolidate the steps such that an entire layer (or set of layers) is treated as a single thermal event. Notwithstanding these challenges, many reports of such simulations were already available at the time of writing (e.g., Promoppatum et al. 2017). Moreover, validation is also being done at various scales demonstrating that predicted thermal histories are feasible (e.g., Beuth and Klingbeil 2001). Residual stress has been shown to be predictable, but thermal distortion is considered to be more challenging for metals because the bed temperature is typically low compared to the melting point (e.g., Szost et al. 2016; Mukherjee et al. 2016). This is in contrast to polymer printing with powders where the effective melting temperature is low in relation to the bed temperature such that thermal stress and distortion are less serious.

## 2   Synchrotron Capabilities

As stated above, the main capabilities that are distinctive to synchrotron-based X-rays are micro-tomography (CT), wide-angle X-ray scattering (WAXS), small-angle X-ray scattering (SAXS), high-energy diffraction microscopy (HEDM), residual stress (RS), and dynamic X-ray radiography (DXR). We now briefly review their respective capabilities to inform microstructure development in additive manufactured materials.

### 2.1   X-Ray Micro-tomography (CT)

Computed tomography at the micron- and nanoscales is a widely used method for the nondestructive characterization of the internal structure of materials. At coarser scales, neutrons (e.g., Cao et al. 2016) and muons (Muon Tomography 2017) are also useful although the author is not aware of an application as yet to AM. It is most effective for large density contrast since it mostly depends on the Beer–Lambert law to measure variations in absorption, aided by phase contrast to detect sharp changes in density (Gursoy et al. 2014). The primary application to AM is for measuring porosity where pores constitute an important defect in materials used in structural applications (Hudák et al. 2016; Eylon and Strope 1979; Scarlett et al. 2016a, b). A pore is a stress riser under tensile load and, above a size of about 10 $\mu$, may act as the source of a fatigue crack as has been documented in many reports on additively manufactured materials (e.g., Leuders et al. 2015). Fatigue is a classic extreme value problem in the sense that the originating defect can be traced to the largest defect available in the material, for which there is a large literature (Weibull 1951). Here again, synchrotron tomography is useful for providing datasets that can then be used to perform simulations of the mechanical response at the relevant length scale (Cunningham et al. 2017b; Kantzos et al. 2018). Figure 1 shows two views of a set
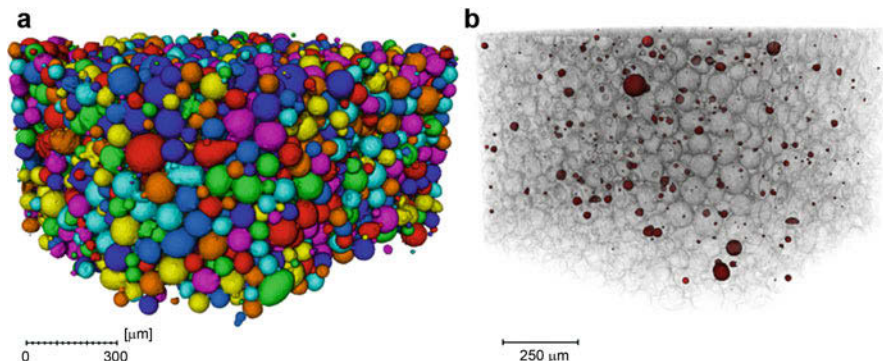
**Fig. 1** (**a**) Tomograph of packed Ti-6Al-4V particles with randomly assigned colors. (**b**) Tomograph of packed Ti-6Al-4V particles (gray outlines) as in (**a**) with voids (pores) colored red for emphasis



**Fig. 2** Schematic diagram of μSXCT setup at 2-BM beamline at Argonne National Lab's Advanced Photon Source
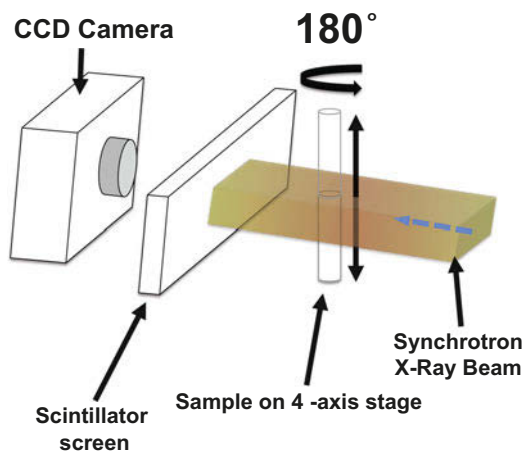
of powder particles of gas atomized Ti-6Al-4V, which is a material commonly used in the aerospace and medical industries; the average size is about 60 μm, which is typical of powders used in laser or electron beam powder bed machines. Panel (a) shows the exterior surface of each particle with randomly assigned colors for contrast; panel (b) shows each void or pore (inside a particle) colored dark red for emphasis. With such fine powders, the sub-micrometer resolution available for high-Z materials with synchrotron radiation is particularly helpful for resolving particles across the full range of size.

Although X-ray micro-CT can be readily performed with laboratory systems, synchrotron-based CT provides better resolution when the sample cross section fits within the beam. Since the latter is typically about 2 mm across, the sample size is quite limited; Fig. 2 provides a schematic view where it is important that

the synchrotron X-ray beam is parallel, by contrast with most lab-based systems that use a diverging beam to be able to illuminate larger samples. Nevertheless, synchrotron CT has shown that porosity in metal powder bed materials has at least main sources. These are (a) a residual porosity inherited from the powders used (Cunningham et al. 2017a), (b) lack-of-fusion porosity that occurs when the melt pools do not overlap sufficiently and some locations are never melted (Tang et al. 2017), and (c) keyhole porosity that arises from excessive penetration by the laser or electron beam (Cunningham et al. 2017b). This parsing of porosity development into distinct regimes contrasts with the more typical approach of relating it to energy density, which is the energy deposited per unit length of weld bead: the energy density is a continuous variable, but recent work supports the existence of thresholds for the number density of pores. The latter also is consistent with the concept of process windows in power-velocity space (Vasinonta et al. 2006), which suggest that there is a range of power-velocity combinations for any given machine and material that yield good-quality builds. The process window is typically an elongated patch whose long dimension corresponds to a particular ratio of power to velocity (Montgomery et al. 2015). Finally, porosity in metal powders has been shown to be ubiquitous (Cunningham et al. 2017b) and in gas atomization at least is a direct consequence of the impingement of high-velocity gas jets that break up the liquid metal stream into fine droplets.

One of the attractive features of AM is the ability to print arbitrarily connected materials over a wide range of densities relative to the material of construction. In powder bed AM, the extensive freedom implied in this approach is constrained by the requirement for a minimum cross section and the difficulty inherent in overhangs, i.e., printing solid material on top of a significant area of unmelted powder. Despite these limitations, numerous results have been published on lattice structures, with a number of original approaches to optimization of the structures (e.g., Calignano 2014). In a lattice, the individual struts in a lattice vary in cross section and have partially melted powder particles attached to their surfaces. The shrinkage associated with solidification and subsequent thermal contraction also results in distortions of the product relative to the original design (Dunbar et al. 2016). As is well-known, the extent of this distortion depends on preheat, the so-called support structure and many other factors. Higher preheat decreases the extent of thermal contraction. Support structures both allow overhanging sections to be printed and attach the part to the baseplate. Such attachment means that the part can be heat treated on the baseplate for stress relief via creep prior to cutting through the support structure to separate it. Synchrotron CT has also proven to be useful for measuring the quality of lattices because it can directly measure the entire 3D structure. Bormann et al. (2013) published an elegant approach in which they continually acquired radiographs as they heated a lattice structure printed in NiTi used as a scaffold for tissue ingrowth. The authors found substantial deviations of the scaffold from the intended design with more locations exhibiting excess material than vice versa. As Khairallah et al. (2016) have pointed out, this is unsurprising because the melting pool tends to pull particles in at its periphery and partially melted particles are likely to be outside the target melting volume.

Brun et al. (2013) reported a similar investigation for polycaprolactone scaffolds, including a comparison between a conventional fabrication method and 3D printed materials.

Carlton et al. (2016) used X-ray micro-CT with in situ tensile testing to investigate damage and fracture in 3D printed stainless steel. They found that the porosity distribution could have a major effect in the sense that specimens with large and inhomogeneous pore distributions exhibited fractures that started from the existing voids. In specimens with low void content, however, the fracture behavior was unaffected by homogeneous distributions of small pores. Sandgren et al. (2016) used micro-CT to study fatigue crack growth in Ti-6Al-4V that had been printed in a powder-feed machine.

## 2.2    Wide-Angle X-ray Scattering (WAXS)

Cakmak et al. (2016) used wide-angle X-ray scattering at the APS to measure texture in electron beam-printed IN718, a Ni-based superalloy, which exhibited strong fiber textures parallel to the build direction (BD). The strong fiber textures with <100>//BD were consistent with the strongly columnar microstructures observed. Although synchrotron-based X-rays have some advantages for texture determination because of its penetrating power, a combination of EBSD and lab-based X-ray pole figures are effective for most investigations.

## 2.3    Small-Angle X-ray Scattering (SAXS)

One notable advantage of synchrotrons is the availability of high-energy X-rays with high intensities, which allows deep penetration and therefore substantial volumes to be interrogated. Many alloys are heat treated, most often to develop increased strength or to arrive at an optimum combination of strength and toughness. Additively manufactured metals often exhibit heat treatment responses that differ from those established for conventionally processed material. As an example, the Ni-based alloys IN718 and IN625 are known to be hardened by such phases as $\gamma'$ and Laves, with carbides and the $\delta$ phase being more important for grain size control. Zhang et al. (2017) used SAXS to study the homogenization kinetics for the major alloying elements, i.e., Ni, Cr, Nb, and Mo. Using the characteristic streaking patterns, they determined that most of the segregation is confined to a region within 6 nm of the center of each dendrite. Correlation analysis between the successive SAXS images provided quantitative information about the kinetics of the diffusion-controlled homogenization process. Similarly, Xue et al. (2015) used SAXS to detect precipitate formation under high cooling rate conditions in Ti-48Al and an an Al-Cu-Mg alloy. The hardening response in IN625 was further investigated by Idell et al. (2016) and Lass et al. (2017) who used conventional methods to determine that the $\delta$ phase was dominant from the outset. Rather than appearing heterogeneously at grain boundaries, the $\delta$ phase precipitated throughout the material in the form of

disc-shaped plates with a well-defined orientation relationship with the matrix. Idell et al. (2015) performed a similar study on IN718 and noted that synchrotron-based SAXS was again useful for measuring the progress of the precipitation process because of the large flux and high energy of the X-rays.

## 2.4    High-Energy Diffraction Microscopy (HEDM)

There are several diffraction-based techniques that use synchrotron X-rays to map polycrystalline microstructures at various levels of resolution (Rollett and Barmak 2015). high-energy diffraction microscopy (HEDM) is one class of those techniques that itself exists in two major variants. One of these is known as near-field HEDM (nf-HEDM) because the detector is placed relatively close to the sample such that the locations of diffraction spots are as sensitive to the point of origin of the diffracted beam in the sample as it is to the diffraction order (i.e., which crystallographic plane). The sample is rotated in front of the beam to excite a large (at least 20) number of beams from each location; the beam is generally planar, to limit the number of spots in each view. A simulated annealing algorithm sifts through the potential orientations in the material to match as large a fraction of the spots as possible and index the orientation at each point in a regular grid (Li and Suter 2013). The far-field variant of HEDM (ff-HEDM) acquires data from a detector whose position is far enough from the sample that Bragg rings are obvious, which greatly facilitates indexation of orientations, but close enough such that deviations of individual spots from the standard position give information on both the center of mass of each diffracting grain and the (average) elastic strain (Bernier et al. 2011). Again, the sample is rotated in front of the beam so that the spot position varies as each grain rotates around the rotation axis. The reconstruction provides information on each grain in a similarly nondestructive manner to nf-HEDM with additional information about the state of elastic strain but not a 3D orientation map (by analogy to EBSD (Schwartz et al. 2000).

Quoting from Rollett and Barmak (2015), there now exists a way to reconstruct "the orientation map of a material from diffraction data acquired with high energy x-rays that can penetrate the full cross-section of a sample up to about 1 millimeter thickness. This means working with beams with energies between 10 and 100 keV that may be focused or parallel, monochromated or broad (white) spectrum and so on. One essential difference between such a technique and other orientation mapping methods, e.g., EBSD (Schwartz et al. 2000), is that illumination of a volume with many grains means it is infeasible to obtain an individual diffraction pattern for each point. Instead, one must infer the orientation of each point by fitting to the entire set of points in real space that have contributed to a (large) set of diffraction patterns. To set the scene for this method, it is convenient to contrast "far-field" from "near-field" approaches. In far-field synchrotron microscopy, the detector is placed of order 1 m away from the sample (for energies in the range 20–100 keV) such that several Bragg rings are imaged. Orientation mapping in 3D that distinguishes grain shapes from merely centers of mass, however, requires the

near-field approach in which the detector (or effective imaging surface) is placed at distances of a few millimeters from the specimen, such that many diffraction peaks are acquired up to a high order of reflection. A flat monochromatic beam illuminates the entire cross section of a sample, which is generally a wire no more than 1 mm in diameter. For both ff- and nf- approaches, the sample is rotated in 1° (or less) steps through a range of 180°, and a diffractogram is acquired at each rotation angle integrated over the interval to ensure that all points in the illuminated volume contribute equally to the dataset as a whole. This data acquisition process results in a large set of diffractograms for each layer, all of which contain information from (potentially) all locations. Note that no specimen preparation is required."

In ff-HEDM, precession of diffraction spots around the rotation axis provides information on the center of mass of each diffracting grain. The deviation of spots from their nominal locations on each ring provides information on elastic strain. The reconstruction of an orientation map consists in searching orientation space for all locations in the illuminated layer simultaneously while using the match between simulated diffraction spots and measured spots as the measure of completeness or confidence in the result (Bernier et al. 2011; Sharma et al. 2012a, b). For the nf-HEDM approach, Li and Suter (2013) built upon multiple previous works to devise an efficient, multiscale method for performing reconstructions and arriving at a 3D orientation map as a stack of layerwise maps. This provides direct information on grain shape but does not resolve elastic strain; the typical resolution is 0.1° in orientation with 2 μm in each layer and 4 μm in the stacking direction. These limits are determined by the detector resolution and beam thickness, respectively. A variant of HEDM can be described as very-far-field HEDM (Lienert et al. 2011), in which a single diffraction spot is imaged on a detector placed a few meters from the sample such that its shape can be monitored as a function of time, temperature, loading, etc.

Recently, researchers in this area have moved to combine the far-field (ff-HEDM) and near-field (nf-HEDM) variants of HEDM. As one example, ff-HEDM (Bernier et al. 2011) can be performed on a 3D volume, which provides a list of grains with their centers of mass and orientations (as well as elastic strain). This list can then be used to "seed" the nf-HEDM reconstruction, which saves substantial computation time because the search in orientation space is otherwise a very time-consuming process. The nf-HEDM reconstruction then provides a more accurate spatial map of the polycrystal microstructure. Such a map provides the basis for instantiating full-field micro-mechanical simulations. Turner et al. (2017) used the data from such a combined experiment (Turner et al. 2016) to instantiate full-field finite element simulations. Comparison of the calculated elastic strains with the measured values showed good agreement. However, they were not able to incorporate the residual elastic strain present in the undeformed material. Consequently, they had to compare changes in strain values using the initial values as the reference point. Following previous efforts to incorporate residual stresses as eigenstrains (e.g., Salvati et al. 2017), Pokharel and Lebensohn (2017) demonstrated that it is possible to approximate the initial (residual) strain at the grain scale via an eigenstrain calculation based on Eshelby, which is important as most measurements of elastic strain on annealed polycrystalline materials have revealed

appreciable levels of residual strain, e.g., Oddershede et al. (2010). Nevertheless, the eigenstrain calculation (Pokharel and Lebensohn 2017) is only the first step toward accommodating such residual strain (stress) conditions into simulations of polycrystal deformation as well as the challenges of understanding their origins. This suggests that modeling of the materials processing that precedes the production of annealed material may be helpful for testing the various hypotheses that have been put forward.

## 2.5    Residual Stress (RS)

The extensive literature on residual stress makes it clear that it is a significant issue in welding. The main reason is that the deformation induced by the heating around the melt pool is not fully reversible. Above a certain temperature, nearly all materials relax such that, at the maximum temperature in the heat-affected zone of a weld, the strain is low (relaxed) and the subsequent cooling results in thermal contraction. The contraction is an eigenstrain that generates stress once the material has cooled below the stress relaxation temperature, i.e., below the point where rapid creep occurs. Of course, the magnitude of the residual stress and its tensorial character depends on the geometry and the extent to which the material around the weld is constrained, but this provides the basic picture. In most cases where deep penetration (with a keyhole) is not being used. The melt pool is semicircular in cross section, and most of the variation in stress is related to the length to width ratio (Gratzke et al. 1992). As already mentioned, additive manufacturing technologies that use selective melting (with lasers or electron beams) are essentially micro-welding processes that focus on melting and consolidation of powders, rather than joining. Therefore, the issues around residual stress in AM have the same physical basis as in welding. Significant contributions abound such as Mach et al. (2017) who have demonstrated that synchrotron radiation can be effectively used to measure the spatial variation in the full elastic strain tensor (from which stress can be derived).

The application to additive manufacturing arises naturally through the fact that SLM is, to some approximation, a process that deposits successive solidified layers of order 50 $\mu$m thick. The shrinkage associated with each layer results in a residual stress that tends to impose a biaxial shrinkage on the part. As mentioned elsewhere, the AM machine manufacturers typically recommend the use of preheat which is commonly much higher in electron-beam machines than in SLM. Whitesell et al. (2016) describe the use of high energy x-rays for residual stress measurement using a laboratory source. Park and Okasinski (2017) discuss techniques for measuring the elastic strains in a spatially resolved manner with synchrotron radiation. The basic idea consists of using a slit to select in which part of the sample a diffracted beam originates. With conical slits, a region of order 150 $\mu$m long can be isolated within the path traversed by the input beam. Park and Okasinski (2017) report on the use of spiral slits that give better performance despite only allowing small portions of each Bragg ring to reach the detector. Nevertheless, the spiral slits rotate to capture an entire Bragg ring. The authors point out that the spiral slits are particularly

helpful for materials with crystal symmetry lower than cubic because each set of conical slits must be constructed for the specific Bragg angles of interest. The experiment described in this paper was on a test article, however, and no application for additively manufactured material has yet been reported.

Mishurova et al. (2017) describe a typical test piece consisting of an arch or bridge that was $8 \times 10 \times 20$ mm in height $\times$ depth $\times$ length. The part was printed from Ti-6Al-4V powder in an SLM machine, and measurements were made at several locations, which meant that the spatial resolution that could be attained was not explored although the closest spacing was 0.5 mm between beam locations. The measurements were performed at a synchrotron in Germany with a white beam in the energy range 10–150 keV. Their main result was that low energy densities, implying higher speed and lower energy, resulted in significantly higher stresses with maximum elastic strains approaching $2.10^{-3}$, which corresponds to stresses a bit under 1 GPa. Not surprisingly, the residual strain state changed after the part was removed from the baseplate, which reflects the removal of the constraint imposed by the latter.

There is significant literature regarding modeling of residual stress. In its more direct form, the thermal history is simulated by the passage of a moving heat source, often simplified to that of a moving point source. This latter simplification has the advantage that the analytical Rosenthal solution is available (Rosenthal 1941), which allows for straightforward calibration against experimental data for melt pool size (Promoppatum et al. 2018). Standard algorithms are applicable, and a wide variety of commercially supported finite element codes are being used. Simulating thermal distortion requires solution of the thermomechanical problem, which in turn means that the mechanical and thermal properties of the material must be known with sufficient accuracy up to the melting point. Heigel et al. (2015) mention the need for a "measurement-based convection model" in order to obtain accurate results. Since the development of thermal strains depends, as mentioned above, on the variation in strength over temperature, another challenge is the lack of detailed data on mechanical properties such as creep strength, Poisson ratio, modulus, etc. close to the melting point. The assumption of a point heat source is unlikely to be accurate, given that there is strong evidence for deep penetration (keyhole) conditions, e.g., Trapp et al. (2017), which results in a large aspect ratio in the melt pool and, effectively, a line heat source. There are, of course, many papers being published that model the powder bed modeling process with varying degrees of completeness. Lindgren et al. (2016), for example, provide a relatively complete such example that includes residual stress calculation as well as material properties.

## 2.6    Dynamic X-Ray Radiography (DXR)

All processes acquire new meaning when they can be visualized directly. The high-speed, highly localized melting and refreezing associated with selective laser or electron beam melting used in many 3D printing machines is no exception. Many valuable experiments have been conducted with high-speed optical imaging

**Fig. 3** Illustration of the arrangement for passing a high-energy X-ray beam through a sample of order 1 mm thick with or without powder placed on top for experiments to investigate selective laser melting (Zhao et al. 2017)
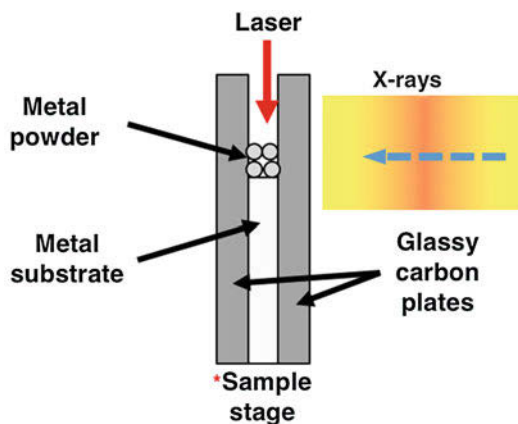


systems that look down on the process (e.g., Furumoto et al. 2013). Nevertheless, significant assumptions must be made about the temperature fields where pyrometry is attempted and filtering of the images is typically required. The advent of direct high-speed radiography as presented by Zhao et al. (2017) promises to transform this area. The sample is confined between transparent carbon plates in a vacuum-capable chamber with windows for the X-rays and a fiber-optic feedthrough for the 520 W laser; the laser system is of the same type as used in SLM machines (Fig. 3). Using a $2 \times 2$ mm pink X-ray beam at beamline 32ID at the APS and a high-speed imaging system, they imaged local spot melting over a range of power levels, with and without powder. The expulsion of powder particles was visualized along with direct measurement of the velocities, thus illustrating the importance of gas flow within SLM machines for limiting the redeposition of particles. The formation of the melt pool and vapor cavities was readily apparent. Figure 4 illustrates the sequence of events (Zhao et al. 2017) as the laser strikes, a melt pool forms, powder particles are ejected, and finally a void is left behind as the elongated melt pool freezes. Such a void is characteristic of end-of-track defects in welding, and it also provides a scenario for the systematic arrays of voids that are sometimes observed in SLM (Groeber et al. 2017). Lately, a scanning system has been added (Parab et al. 2018) that opens up the possibility of studying a wide range of power levels and speeds that would be relevant to the practice of SLM.

There exists a substantial literature on modeling melt pools in welding, laser drilling, and additive manufacturing. Khairallah et al. (2016) and Tan and Shin (2015) are two examples of large-scale multi-physics simulations that include heat flow, fluid flow, and gas flow. Khairallah et al. (2016) specifically address the smaller scale of selective laser melting and note the importance of surface tension (and its temperature dependence) and recoil pressure from evaporation. Tan and Shin (2015) emphasize the somewhat larger scale of welding and show the variety of keyhole shapes that result from variations in power and speed. Not so well-known outside this area is the fact that a keyhole is not necessarily a simple cylindrical depression but adopts more complicated, elongated shapes especially at high traverse speeds.
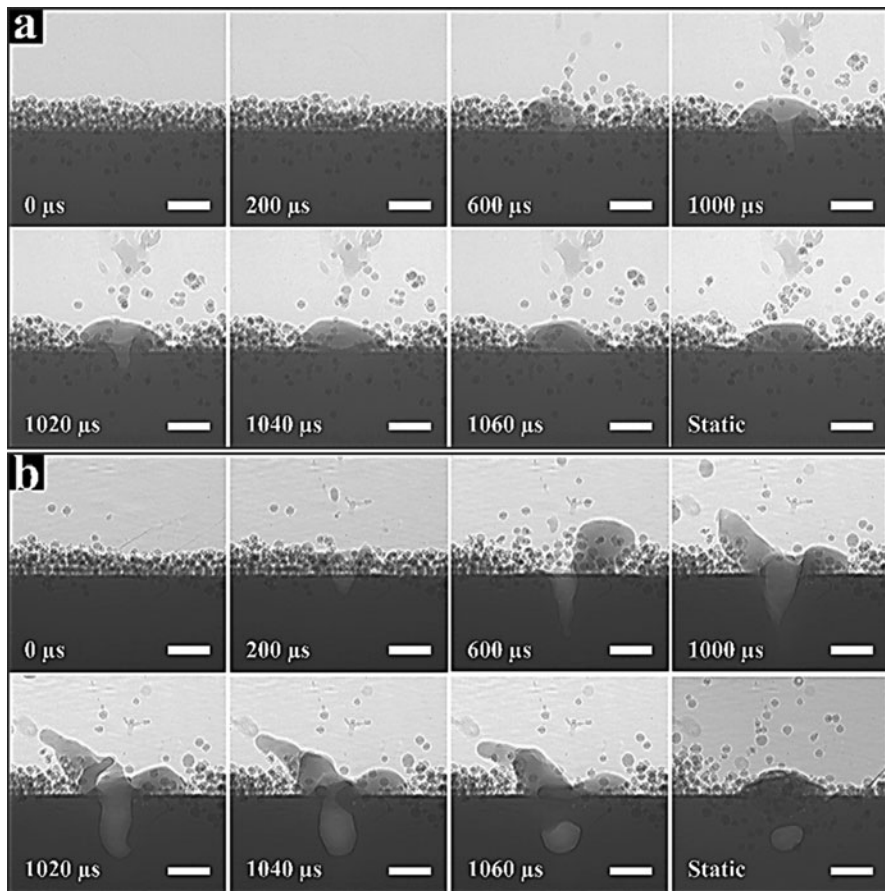
**Fig. 4** Successive frames from an experiment in which a static laser beam impinges on a Ti-6Al-4V sample with powder on top. A melt pool forms, powder particles are ejected from the powder bed, liquid is ejected from the pool, and a void forms as the somewhat elongated melt pool freezes (Zhao et al. 2017)

Notwithstanding the many contributions in this area, there are many opportunities to explore additional aspects of melting at high speeds, particularly with respect to the effects of powders.

## 2.7    Dynamic X-Ray Diffraction

Kenel et al. (2016a, b) performed high-speed diffraction experiments on Ti-48Al at the Paul Scherrer Institute. A custom-designed support was used for the sample
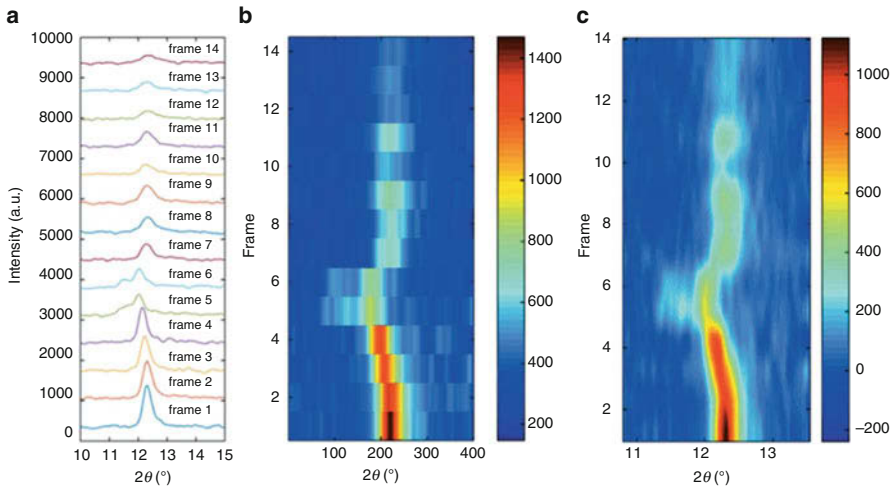
**Fig. 5** Example of results from a high-speed diffraction experiment on a NiTi alloy sample that was subjected to dynamic deformation, reproduced from Sun and Fezzaa (2016). The frame rate was 20 kHz, and the color maps show both the shifts in Bragg angle and the loss of intensity as deformation proceeds. More specifically, (**a**) shows a series of intensity traces, (**b**) is a color map of intensity for frame number versus diffraction angle (binned over the same range as in (**c**)), and (**c**) is a smoothed version of (**b**)

with laser heating to induce melting which approximated the conditions of SLM. They were able to show that solidification proceeded directly to $\alpha$ and $\alpha_2$ with $\gamma$ appearing shortly thereafter. The temporal resolution was of order 10 ms, which was sufficient for the material studied. Zhao et al. (2017) described an even higher-speed experiment with approximately microsecond resolution on Ti-6Al-4V in which they were able to not only measure the melting and resolidification to the high-temperature BCC $\beta$ phase with subsequent transformation to the HCP $\alpha$ phase but also image the progress of the melting and freezing. Thanks to the use of sample dimensions and a focused laser source of the same type used in SLM, they were able to achieve similar cooling rates, etc. to those in actual additive manufacturing machines. Kenel et al. (2017) followed this with a similar high-rate diffraction experiment on high-rate melting and refreezing of Ti-6Al-4V. The lower frame 1 kHz rate and high-sensitivity detector used in this case allowed them to identify the formation of the $\alpha'$ martensitic phase, as well as evolution in the $\beta$ phase. They used the same support structure as in their previous experiment, however, so the conditions were partially representative of AM. The high-speed diffraction results reported by Zhao et al. (2017) were based on developments by Sun and Fezzaa (2016), and Fig. 5 shows in more detail what sort of variations in diffraction can be observed on a 20 kHz timescale.

# 3      Impact on Modeling and Simulation

The foremost need in modeling and simulation is for improvements in the prediction of distortion of parts. This is not a simple task: whereas prediction of thermal histories is reasonably straightforward, computing the actual displacements is not reliable. Temperature fields depend primarily on the heat source(s) and the thermal properties, which are generally well-known. Computing the resulting distortions, however, is much more challenging because additional temperature- and strain rate-dependent properties, such as yield strength, Poisson ratio, and creep rate, must be well defined all the way up to the melting point. Such properties are challenging to measure and many materials are not well characterized.

The high-speed diffraction experiments that are now feasible suggest that a substantial forward modeling effort is needed to simulate the X-ray scattering and move toward a more complete understanding of the dynamic microstructural evolution. Forward modeling of diffraction is practiced in many different areas, but the actual physics that must be included for accurate simulation of any specific experiment depends strongly on the circumstances.

In addition, multiscale modeling of the relations between microstructure and thermomechanical properties, as well as computer simulations of microstructure evolution benefits from the synchrotron experimental studies. For example, X-ray diffraction results such as atomic structure, lattice parameters, and local ordering provide input to atomic and mesoscale modeling of microstructural features. Furthermore, tomography results provide rich validation data for mesoscale simulations of microstructure evolution.

# 4      Conclusions

Most of the ways in which synchrotron X-ray radiation experiments can contribute to advancing our understanding of the additive manufacturing processes have been explored, at least in a preliminary sense. Given the high cooling rates found in additive manufacturing especially with SLM, there is a high likelihood that this area of activity will continue to evolve rapidly and that synchrotron radiation will play a major role in elucidating the scientific challenges associated with additive manufacturing. Wide-angle and small-angle X-ray scattering have been used to measure microstructural evolution and in particular diffusion and precipitation processes. Dynamic X-ray radiography (DXR) has been used to image the melting and refreezing processes directly, with complementary experiments on high-speed diffraction. Notwithstanding the limited volumes that can be scanned, computed tomography (CT) has made numerous contributions thanks to the high spatial resolution available with synchrotron X-rays as well as the relatively high throughput that is feasible. More advanced techniques such as high-energy

diffraction microscopy look likely to make a contribution to 3D characterization of AM materials in the near future. Synchrotron X-ray radiation experiments provide valuable information for setting up and validating atomistic and mesoscale computer simulations of microstructure evolution.

## References

Bernier JV, Barton NR, Lienert U, Miller MP (2011) Far-field high-energy diffraction microscopy: a tool for intergranular orientation and strain analysis. J Strain Anal Eng Des 46:527–547. https://doi.org/10.1177/0309324711405761

Beuth J, Klingbeil N (2001) The role of process variables in laser-based direct metal solid freeform fabrication. JOM 53:36–39. https://doi.org/10.1007/s11837-001-0067-y

Bormann T, de Wild M, Beckmann F, Mueller B (2013) Assessing the morphology of selective laser melted NiTi-scaffolds for a three-dimensional quantification of the one-way shape memory effect. In: Goulbourne NC, Naguib HE (eds) Behavior and mechanics of multifunctional materials and composites. SPIE, Bellingham, p 868914

Brun F, Intranuovo F, Mohammadi S, Domingos M, Favia P, Tromba G (2013) A comparison of 3D poly(epsilon-caprolactone) tissue engineering scaffolds produced with conventional and additive manufacturing techniques by means of quantitative analysis of SR mu-CT images. J Instrum 8:C07001. https://doi.org/10.1088/1748-0221/8/07/C07001

Cakmak E, Kirka MM, Watkins TR, Cooper RC, An K, Choo H, Wu W, Dehoff RR, Babu SS (2016) Microstructural and micromechanical characterization of IN718 theta shaped specimens built with electron beam melting. Acta Mater 108:161–175. https://doi.org/10.1016/j.actamat.2016.02.005

Calignano F (2014) Design optimization of supports for overhanging structures in aluminum and titanium alloys by selective laser melting. Mater Des 64:203–213. https://doi.org/10.1016/j.matdes.2014.07.043

Cao J, Gharghouri MA, Nash P (2016) Finite-element analysis and experimental validation of thermal residual stress and distortion in electron beam additive manufactured Ti-6Al-4V build plates. J Mater Process Technol 237:409–419

Carlton HD, Haboub A, Gallegos GF, Parkinson DY, MacDowell AA (2016) Damage evolution and failure mechanisms in additively manufactured stainless steel. Mater Sci Eng A 651:406–414. https://doi.org/10.1016/j.msea.2015.10.073

Cunningham R, Narra SP, Montgomery C, Beuth J, Rollett A (2017a) Synchrotron-based x-ray microtomography characterization of the effect of processing variables on porosity formation in laser power-bed additive manufacturing of Ti-6Al-4V. JOM 69:479. https://doi.org/10.1007/s11837-016-2234-1

Cunningham R, Nicolas A, Madsen J, Fodran E, Anagnostou E, Sangid MD, Rollett AD (2017b) Analyzing the effects of powder and post-processing on porosity and properties of electron beam melted Ti-6Al-4V. Mater Res Lett 5:516–525. https://doi.org/10.1080/21663831.2017.1340911

Dunbar AJ, Denlinger ER, Heigel J, Michaleris P, Guerrier P, Martukanitz R, Simpson TW (2016) Development of experimental method for in situ distortion and temperature measurements during the laser powder bed fusion additive manufacturing process. Addit Manuf 12:25–30. https://doi.org/10.1016/j.addma.2016.04.007

Eylon D, Strope B (1979) Fatigue crack initiation in Ti-6 wt% Al-4 wt% V castings. J Mater Sci 14:345–353. https://doi.org/10.1007/BF00589825

Furumoto T, Ueda T, Alkahari MR, Hosokawa A (2013) Investigation of laser consolidation process for metal powder by two-color pyrometer and high-speed video camera. CIRP Ann 62:223–226. https://doi.org/10.1016/j.cirp.2013.03.032

Gibson I, Rosen DW, Stucker B (eds) (2010) Additive manufacturing technologies: 3D printing, rapid prototyping, and direct digital manufacturing, 2nd edn. Springer, New York

Gratzke U, Kapadia P, Dowden J, Kroos J, Simon G (1992) Theoretical approach to the humping phenomenon in welding processes. J Phys D 25:1640. https://doi.org/10.1088/0022-3727/25/11/012

Groeber M, Schwalbach E, Donegan S, Chaput K, Butler T, Miller J (2017) Application of characterization, modelling, and analytics towards understanding process-structure linkages in metallic 3D printing. IOP Conf Ser Mater Sci Eng 219:012002

Gürsoy D, De Carlo F, Xiao X, Jacobsen C (2014) TomoPy: a framework for the analysis of synchrotron tomographic data. J Synch Radiat 21:1188–1193

Heigel JC, Michaleris P, Reutzel EW (2015) Thermo-mechanical model development and validation of directed energy deposition additive manufacturing of Ti–6Al–4V. Addit Manuf 5:9–19. https://doi.org/10.1016/j.addma.2014.10.003

Hudák R, Zivák J, Tóth T, Majerník J, Lisý M (2016) Usage of industrial computed tomography for evaluation of custom-made implants. In: Bris R, Majernik J, Pancerz K, Zaitseva E (eds) Applications of computational intelligence in biomedical technology. Springer International Publishing, Switzerland, pp 29–45

Idell Y, Campbell C, Levine L, Zhang F, Olson G, Snyder D (2015) Characterization of nickel based superalloys processed through direct metal laser sintering technique of additive manufacturing. Microsc Microanal 21:465. https://doi.org/10.1017/S1431927615003128

Idell Y, Levine LE, Allen AJ, Zhang F, Campbell CE, Olson G, Gong J, Snyder D, Deutchman H (2016) Unexpected delta-phase formation in additive-manufactured Ni-based superalloy. JOM 68:950–959. https://doi.org/10.1007/s11837-015-1772-2

Kantzos CA, Cunningham RW, Tari V, Rollett AD (2018) Characterization of metal additive manufacturing surfaces using synchrotron x-ray CT and micromechanical modeling. Comput Mech. https://doi.org/10.1007/s00466-017-1531-z

Kenel C, Grolimund D, Fife JL, Samson VA, Van Petegem S, Van Swygenhoven H, Leinenbach C (2016a) Combined in situ synchrotron micro x-ray diffraction and high-speed imaging on rapidly heated and solidified Ti-48Al under additive manufacturing conditions. Scr Mater 114:117–120. https://doi.org/10.1016/j.scriptamat.2015.12.009

Kenel C, Schloth P, Van Petegem S, Fife JL, Grolimund D, Menzel A, Van Swygenhoven H, Leinenbach C (2016b) In situ synchrotron x-ray diffraction and small angle x-ray scattering studies on rapidly heated and cooled Ti-Al and Al-Cu-Mg alloys using laser-based heating. JOM 68:978–984. https://doi.org/10.1007/s11837-015-1774-0

Kenel C, Grolimund D, Li X, Panepucci E, Samson VA, Sanchez DF, Marone F, Leinenbach C (2017) In situ investigation of phase transformations in Ti-6Al-4V under additive manufacturing conditions combining laser melting and high-speed micro-x-ray diffraction. Sci Rep 7:16358. https://doi.org/10.1038/s41598-017-16760-0

Khairallah SA, Anderson AT, Rubenchik A, King WE (2016) Laser powder-bed fusion additive manufacturing: physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones. Acta Mater 108:36–45

Lass EA, Stoudt MR, Williams ME, Katz MB, Levine LE, Phan TQ, Gnaeupel-Herold TH, Ng DS (2017) Formation of the Ni3Nb delta-phase in stress-relieved inconel 625 produced via laser powder-bed fusion additive manufacturing. Metall Mater Trans A 48:5547–5558. https://doi.org/10.1007/s11661-017-4304-6

Leuders S, Vollmer M, Brenne F, Tröster T, Niendorf T (2015) Fatigue strength prediction for titanium alloy TiAl6V4 manufactured by selective laser melting. Metall Mater Trans A 46:3816–3823

Li SF, Suter RM (2013) Adaptive reconstruction method for three-dimensional orientation imaging. J Appl Crystallogr 46:512–524. https://doi.org/10.1107/S0021889813005268

Lienert U, Li SF, Hefferan CM, Lind J, Suter RM, Bernier JV, Barton NR, Brandes MC, Mills MJ, Miller MP, Jakobsen B, Pantleon W (2011) High-energy diffraction microscopy at the advanced photon source. JOM 63:70–77

Lindgren L, Lundbäck A, Fisk M, Pederson R, Andersson J (2016) Simulation of additive manufacturing using coupled constitutive and microstructure models. Addit Manuf 12:144–158. https://doi.org/10.1016/j.addma.2016.05.005

Lundbäck A, Pederson R, Colliander MH, Brice C, Steuwer A, Heralic A, Buslaps T, Lindgren L (2016) Modeling and experimental measurement with synchrotron radiation of residual stresses in laser metal deposited Ti-6Al-4V. In: Proceedings of the 13th world conference on titanium. pp 1279–1282

Mach JC, Budrow CJ, Pagan DC, Ruff JPC, Park J, Okasinski J, Beaudoin AJ, Miller MP (2017) Validating a model for welding induced residual stress using high-energy x-ray diffraction. JOM 69:893–899. https://doi.org/10.1007/s11837-017-2298-6

Mishurova T, Cabeza S, Artzt K, Haubrich J, Klaus M, Genzel C, Requena G, Bruno G (2017) An assessment of subsurface residual stress analysis in SLM Ti-6Al-4V. Materials 10:348. https://doi.org/10.3390/ma10040348

Montgomery C, Beuth J, Sheridan L, Klingbeil N (2015) Process mapping of inconel 625 in laser powder bed additive manufacturing. In: Bourell D (ed) Solid freeform fabrication symposium. Austin, TX, pp 1195–1204

Mukherjee T, Zuback J, De A, DebRoy T (2016) Printability of alloys for additive manufacturing. Sci Rep 6:19717. https://doi.org/10.1038/srep19717

Muon Tomography (2017) Available via https://en.wikipedia.org/wiki/Muon_tomography. Accessed Feb 2018

Murr LE, Johnson WL (2017) 3D metal droplet printing development and advanced materials additive manufacturing. J Mater Res Technol 6:77–89. https://doi.org/10.1016/j.jmrt.2016.11.002

Oddershede J, Schmidt S, Poulsen HF, Sorensen HO, Wright J, Reimers W (2010) Determining grain resolved stresses in polycrystalline materials using three-dimensional x-ray diffraction. J Appl Crystallogr 43:539–549

Parab ND, Zhao C, Cunningham RW, Escano LI, Fezzaa K, Everhart W, Rollett AD, Chen L, Sun T (2018) Ultrafast x-ray imaging of laser metal additive manufacturing processes. Submitted

Park J, Okasinski J (2017) Non-destructive internal lattice strain measurement using high energy synchrotron radiation. In: Quinn S, Balandraud X (eds) Residual stress, thermomechanics and infrared imaging, hybrid techniques and inverse problems, vol 9. Springer Nature, Switzerland, pp 121–126

Pokharel R, Lebensohn RA (2017) Instantiation of crystal plasticity simulations for micromechanical modelling with direct input from microstructural data collected at light sources. Scr Mater 132:73–77. https://doi.org/10.1016/j.scriptamat.2017.01.025

Promoppatum P, Yao S, Pistorius PC, Rollett AD (2017) A comprehensive comparison of the analytical and numerical prediction of the thermal history and solidification microstructure of inconel 718 products made by laser powder-bed fusion. Engineering 3:685–694. https://doi.org/10.1016/J.ENG.2017.05.023

Promoppatum P, Yao S, Pistorius PC, Rollett AD, Coutts PJ, Lia F, Martukanitz R (2018) Numerical modeling and experimental validation of thermal history and microstructure for additive manufacturing of an inconel 718 product. Prog Addit Manuf 3:1–18. https://doi.org/10.1007/s40964-018-0039-1

Ren YQ, King PC, Yang YS, Xiao TQ, Chu C, Gulizia S, Murphy AB (2017) Characterization of heat treatment-induced pore structure changes in cold-sprayed titanium. Mater Charact 132:69–75. https://doi.org/10.1016/j.matchar.2017.08.006

Rollett AD, Barmak K (2015) Orientation mapping. In: Laughlin DE, Hono K (eds) Physical metallurgy, 5th edn. Elsevier, London, pp 1113–1142

Rosenthal D (1941) Mathematical theory of heat distribution during welding and cutting. Weld J 20:220s–234s

Salvati E, Lunt AJG, Ying S, Sui T, Zhang HJ, Heason C, Baxter G, Korsunsky AM (2017) Eigenstrain reconstruction of residual strains in an additively manufactured and shot peened nickel superalloy compressor blade. Comp Methods Appl Mech Eng 320:335–351. https://doi.org/10.1016/j.cma.2017.03.005

Sandgren HR, Zhai Y, Lados DA, Shade PA, Schuren JC, Groeber MA, Kenesei P, Gavras AG (2016) Characterization of fatigue crack growth behavior in LENS fabricated Ti-6Al-4V using high-energy synchrotron x-ray microtomography. Addit Manuf 12:132–141. https://doi.org/10.1016/j.addma.2016.09.002

Scarlett NVY, Tyson P, Fraser D, Mayo S, Maksimenko A (2016a) Synchrotron x-ray CT characterization of titanium parts fabricated by additive manufacturing. Part I. Morphology. J Synchrotron Radiat 23:1006–1014. https://doi.org/10.1107/S1600577516007359

Scarlett NVY, Tyson P, Fraser D, Mayo S, Maksimenko A (2016b) Synchrotron x-ray CT characterization of titanium parts fabricated by additive manufacturing. Part II. Defects. J Synchrotron Radiat 23:1015–1023. https://doi.org/10.1107/S1600577516008018

Schwartz AJ, Kumar M, Adams BL (eds) (2000) Electron backscatter diffraction in materials science, 2nd edn. Kluwer, New York

Sharma H, Huizenga RM, Offerman SE (2012a) A fast methodology to determine the characteristics of thousands of grains using three-dimensional x-ray diffraction. II. Volume, centre-of-mass position, crystallographic orientation and strain state of grains. J Appl Crystallogr 45:705–718. https://doi.org/10.1107/S0021889812025599

Sharma H, Huizenga RM, Offerman SE (2012b) A fast methodology to determine the characteristics of thousands of grains using three-dimensional x-ray diffraction. I. Overlapping diffraction peaks and parameters of the experimental setup. J Appl Crystallogr 45:693–704. https://doi.org/10.1107/S0021889812025563

Sun T, Fezzaa K (2016) HiSPoD: a program for high-speed polychromatic x-ray diffraction experiments and data analysis on polycrystalline samples. J Synchrotron Radiat 23:1046–1053. https://doi.org/10.1107/S1600577516005804

Szost BA, Terzi S, Martina F, Boisselier D, Prytuliak A, Pirling T, Hofmann M, Jarvis DJ (2016) A comparative study of additive manufacturing techniques: residual stress and microstructural analysis of CLAD and WAAM printed Ti–6Al–4V components. Mater Des 89:559–567. https://doi.org/10.1016/j.matdes.2015.09.115

Tan W, Shin YC (2015) Multi-scale modeling of solidification and microstructure development in laser keyhole welding process for austenitic stainless steel. Comp Mater Sci 98:446–458. https://doi.org/10.1016/j.commatsci.2014.10.063

Tang M, Pistorius PC, Beuth JL (2017) Prediction of lack-of-fusion porosity for powder bed fusion. Addit Manuf 14:39–48. https://doi.org/10.1016/j.addma.2016.12.001

Trapp J, Rubenchik AM, Guss G, Matthews MJ (2017) In situ absorptivity measurements of metallic powders during laser powder-bed fusion additive manufacturing. Appl Mater Today 9:341–349. https://doi.org/10.1016/j.apmt.2017.08.006

Turner TJ, Shade PA, Bernier JV, Li SF, Schuren JC, Lind J, Lienert U, Kenesei P, Suter RM, Blank B et al (2016) Combined near-and far-field high-energy diffraction microscopy dataset for Ti-7Al tensile specimen elastically loaded in situ. Integr Mater Manuf Innov 5:5. https://doi.org/10.1186/s4019

Turner TJ, Shade PA, Bernier JV, Li SF, Schuren JC, Kenesei P, Suter RM, Almer J (2017) Crystal plasticity model validation using combined high-energy diffraction microscopy data for a Ti-7Al specimen. Metall Mater Trans A 48:627–647. https://doi.org/10.1007/s11661-016-3868-x

Vasinonta A, Griffith M, Beuth JL (2006) Process maps for predicting residual stress and melt pool size in the laser-based fabrication of thin-walled structures. J Manuf Sci Eng 129:101–109

Weibull W (1951) A statistical distribution of wide applicability. ASME J Appl Mech 18:293–297

Whitesell R, McKenna A, Wendt S, Gray J (2016) Volumetric measurement of residual stress using high energy x-ray diffraction. In: Chimenti DE, Bond LJ (eds) 42nd annual review of progress in quantitative nondestructive evaluation: incorporating the 6th European-American workshop on reliability of NDE. AIP, College Park, p 110013

Wohlers T (ed) (2017) Wohlers report 2017. Wohlers Associates, Frisco

Xue J, Zhang A, Li Y, Qian D, Wan J, Qi B, Tamura N, Song Z, Chen K (2015) A synchrotron study of microstructure gradient in laser additively formed epitaxial Ni-based superalloy. Sci Rep 5:14903. https://doi.org/10.1038/srep14903

Zhang F, Levine LE, Allen AJ, Campbell CE, Lass EA, Cheruvathur S, Stoudt MR, Williams ME, Idell Y (2017) Homogenization kinetics of a nickel-based superalloy produced by powder bed fusion laser sintering. Scr Mater 131:98–102. https://doi.org/10.1016/j.scriptamat.2016.12.037

Zhao C, Fezzaa K, Cunningham RW, Wen H, Carlo F, Chen L, Rollett AD, Sun T (2017) Real-time monitoring of laser powder bed fusion process using high-speed x-ray imaging and diffraction. Sci Rep 7:3602. https://doi.org/10.1038/s41598-017-03761-2

# Computational Modeling of Morphology Evolution in Metal-Based Battery Electrodes

# 54

Venkat Srinivasan, Kenneth Higa, Pallab Barai, and Yuanyuan Xie

## Contents

V. Srinivasan (✉) · P. Barai
Argonne National Laboratory, Lemont, IL, USA
e-mail: vsrinivasan@anl.gov; baraip@anl.gov

K. Higa
Lawrence Berkeley National Laboratory, Berkeley, CA, USA
e-mail: khiga@lbl.gov

Y. Xie
California State University, Fresno, CA, USA
e-mail: yyxie.cn@gmail.com

**Abstract**

Superior energy and power density, low toxicity, and enhanced shelf life have contributed to the popularity of lithium ion batteries as energy storage devices in the electronics and automobile industries. However, next-generation lithium ion batteries will require even higher energy densities to meet ever-increasing demands for longer battery life. Owing to its extremely high theoretical specific capacity (approximately ten times larger than that of conventional anode materials) and low electrochemical reduction potential ($-3.04$ V with respect to $H/H^+$ reference electrode), lithium metal is a highly attractive candidate as an anode material for next-generation lithium ion batteries. However, challenges such as dendrite growth, which can lead to short circuits or capacity loss from electrical isolation of growths, have prevented widespread commercial use of lithium metal electrodes. Successful commercialization will require stabilization of lithium deposition. Devising strategies to achieve this goal will require an understanding of the fundamental mechanisms that govern electrochemical deposition and dendrite propagation and which span multiple length scales. Building on experimental observations, several mathematical models have been developed to evaluate the roles of a variety of physical phenomena (such as electrochemical reaction, diffusion, migration, mechanical stress and strain, and surface tension) in lithium deposition processes. The present chapter provides an overview of these approaches for modeling lithium deposition and dendrite growth, summarizes their findings, and discusses remaining questions and future directions for dendrite modeling.

# 1 Introduction

For the development of a sustainable economy and to minimize global warming and solve future energy problems, significant emphasis is being given on the implementation of renewable energy sources. Energy storage devices are of utmost importance for achieving this reality. The intermittent renewable energy sources require high energy density devices for storing excess energy produced at times of low demand and supplying stored energy during times of peak consumption. Appropriate storage devices are particularly needed to power automobiles with nonfossil energy sources.

The scientific community has known about electrochemical cells, in which energy can be stored in the form of chemical energy and extracted as electrical energy (Blomgren 2017), for almost one-and-a-half centuries. However, their usage

in commercial devices became popular only around 50 years ago with the advent of rechargeable lead-acid batteries (Rahn 2013; Blomgren 2017). In the last quarter of the twentieth century, the nickel-cadmium and nickel-metal hydride batteries came to the market. These offered higher energy density than lead-acid based chemistries (Rahn 2013), and their lower weight and toxicity made them good options for hand-held electronic devices – the first-generation mobile phones and laptop computers used these nickel-metal hydride-based batteries. Towards the end of the last century, Sony revolutionized the energy storage industry with the commercialization of the first battery based on lithium ion chemistry (Blomgren 2017). It not only contained significantly higher energy and power density, but also demonstrated higher efficiency than the lead-acid and nickel-metal hydride based chemistries (Rahn 2013). The enhanced shelf life, higher capacity retention over multiple cycles, and negligible memory effect have made the lithium ion chemistry a significantly superior option in the electronics industry. However, in the automobile industry, the commercially available lithium ion batteries are far behind gasoline engines in terms of energy and power density (Gallagher et al. 2014). Significant improvements over the present-day lithium ion batteries are required for successful electrification of the automobile industry (Gallagher et al. 2014).

## 1.1 Brief History and Future Directions of Lithium Ion Battery Technology

Sony Corporation first commercialized lithium ion batteries in 1991, using a carbon-based anode, lithium cobalt oxide cathode, and organic liquid electrolytes solvents (Blomgren 2017). Both the cathode and anode material store lithium through an intercalation mechanism, and the ions shuttle between the two electrodes during the charge and discharge process. Carbon was used as the anode material due to its low electrochemical potential (with respect to $Li^+/Li$) and ability to store lithium without the formation of dendrites. In the first commercially successful version, the lithium ion chemistry was able to demonstrate an energy density of 200 Wh/l (Blomgren 2017). This was sufficient for powering electronic devices, but not good enough for running electric vehicles.

To reduce the cost and toxicity associated with cobalt-based cathode materials, several other transition metals have been investigated as cathode materials. For example, nickel-rich cathodes are being used widely for their high capacity, and manganese-rich cathodes have also been investigated for their high stability. Presently, certain combinations of nickel, manganese, and cobalt ($Ni_xMn_yCo_{(1-x-y)}$) are used as lithium ion battery cathodes to combine their intrinsic properties of high capacity, enhanced electrochemical stability, and large electronic conductivity, respectively (Myung et al. 2017). These advanced cathode materials are capable of demonstrating energy densities in the range of 400 Wh/l, which are sufficient to meet the needs of short range electric vehicles (Blomgren 2017; Myung et al. 2017).

In order to overcome the range anxiety associated with the present-day electric vehicles, the energy density of existing lithium ion batteries must be improved. It can be achieved in several ways, some of which are listed below:

(a) Increasing the window of operation for cell potential. This can be achieved by charging the cell to a higher voltage and extracting some excess lithium from the cathode (Myung et al. 2017). This has the disadvantage of disrupting the crystal structure of the cathode material through the removal of oxygen from the lattice sites. This irreversible change leads to high capacity only at the first cycle.
(b) Decreasing the anode potential can also lead to a larger potential window of operation (Xu et al. 2014). Only lithium metal itself demonstrates a lower potential than graphite. Hence, a transition in the anode material, from conventional graphite to lithium metal, is of utmost importance for increasing the energy density of present-day electric vehicles.
(c) Increasing the specific capacity of cathode and anode active materials used in the electrode. Due to their higher specific capacities, nickel-rich NMC and NCA type cathode materials have been investigated thoroughly (Nitta et al. 2015; Myung et al. 2017). From the perspective of the anode, specific capacity can be improved by replacing graphite with either lithium, or silicon/tin (Si/Sn) based active materials (Nitta and Yushin 2014).
(d) Adoption of "beyond lithium ion" chemistries, such as lithium-sulfur- and/or lithium-air-based batteries (Gallagher et al. 2014). These future-generation technologies use lithium metal as the anode material and some unconventional cathode material. For example, lithium-sulfur cells use phase changing sulfur as the cathode that can consume theoretically sixteen electrons during a discharge process. Lithium oxygen/air batteries demonstrate almost zero theoretical weight of the cathode, giving them an excessively high energy density almost comparable to that of gasoline.
(e) Replacement of liquid electrolytes by solid-state ceramics- or polymer-based electrolytes (Kerman et al. 2017). Due to the higher safety associated with nonliquid electrolytes, they might allow the removal of certain extra safety devices required for the liquid electrolytes. This extra space can be used for inserting extra cells and increasing the pack level energy density. Solid-state electrolytes can also help to accelerate the implementation of lithium metal anodes.

## 1.2    Lithium Metal as the Negative Electrode

Most of the above techniques for increasing energy density involve lithium metal as an anode, as it possesses an extremely low electrochemical potential of −3.04 V with respect to hydrogen reference electrode, and theoretical specific capacity of 3870 mAh/g, which is almost an order-of-magnitude higher than the conventional graphite based anodes (375 mAh/g) (Xu et al. 2014; Lin et al. 2017). It should be noted that lithium metal anodes were actively researched in the second half of the

last century for usage in lithium ion batteries, but did not see commercial success due to several issues (Blomgren 2017; Cheng et al. 2017):

(a) Loss of electrolyte due to reaction with the anode surface and formation of a solid electrolyte interface, which leads to significant capacity fade
(b) Formation of dead lithium that does not participate in the electrochemical reaction and can potentially lead to capacity fade if excess lithium is not used
(c) Growth of dendritic protrusions that can penetrate through the separator and short the cell
(d) Thermal runaway due to internal shorting and subsequent explosion of the cell

Nucleation and propagation of dendritic protrusions during lithium deposition are connected to all of these problems. Significant research has been conducted to understand and prevent these phenomena. For example, it was demonstrated that the use of a small amount of hydrofluoric acid within the electrolyte as an additive results in the formation of a mechanically stiff lithium-fluoride layer on top of the metal anode, which can stabilize the metal-electrolyte interface for tens of cycles (Takehara 1997). Later it was hypothesized that nonuniform lithium deposition at the electrode-SEI interface and subsequent rupture of the solid electrolyte interphase layer result in the nucleation of the lithium dendrites (Aurbach 2000). Addition of stiff protective layers along with the separator can prevent growth of dendritic protrusions. Some examples of such protective layers are sulfur and nitrogen co-doped graphene (Shin et al. 2015) and alumina ($Al_2O_3$) deposited on top of lithium (Peng et al. 2016). Covering the lithium surface with polymer networks or carbon nano-spheres can help to minimize the nonuniform lithium deposition or SEI formation, respectively (Zheng et al. 2014; Peng et al. 2016). There have also been several research efforts to alter the transport properties of electrolytes (conductivity, diffusivity, and transference number) in order to prevent the growth of lithium dendrites. For example, plasticizers have been added within polymer electrolytes to improve conductivity near room temperature conditions for stabilizing lithium deposition (Khurana et al. 2014). Nanoparticles have also been added within the liquid electrolyte to immobilize the anions and increase the lithium transference number, which helps to maintain higher anion concentration near the anode (Lu et al. 2012). However, to the best of the authors' knowledge, no conclusive product or strategy has been developed that can ensure stable deposition of lithium for thousands of cycles with minimal capacity fade.

Next-generation polymer or solid-state ceramic electrolytes were not readily available in the earlier efforts, but now make practical lithium metal anodes appear to be more achievable. However, preliminary experimental studies at the lab scale still show the formation of lithium dendrites through these nonliquid electrolytes (Harry et al. 2016; Kerman et al. 2017). Better design of these electrolytes may be able to stabilize the lithium deposition process.

Furthermore, different dendrite morphologies have been observed, varying according to the type of electrolyte, operating conditions, and applied overpotential (Harry et al. 2016; Cheng et al. 2017):

(a) Mossy deposits form at low current densities. This is usually associated with lithium deposition at the base of the dendrite. If the lithium deposition process is charge-transfer controlled, moss-like deposits are expected to form (Bai et al. 2016).

(b) Needle-like protrusions, which evolve during deposition at higher rates, are usually associated with deposition at the tip of the protrusion. In mass transfer limited lithium deposition processes, needle-like dendrites are observed to form (Bai et al. 2016).

(c) Tree-like dendritic growths have also been observed and are usually attributed to the presence of surface inhomogeneity.

(d) Globular dendritic protrusions have also been reported during deposition under the influence of mechanical stress (Harry et al. 2016).

Among all the different possible microstructures, needle-shaped dendrites have the largest potential to penetrate through the separator, get in contact with the cathode, and short the cell (Li et al. 2014; Bai et al. 2016). Hence, even if dendrite growth is unavoidable, it is important to control the lithium microstructure that evolves during the deposition process and possibly operate the cell only under conditions in which it forms only mossy deposits. In order to develop strategies for controlling the lithium deposition process, it is important to understand the overall lithium dendrite nucleation and growth mechanisms.

## 1.3   Mathematical Models for Lithium Dendrite Nucleation and Growth

Extensive research on electrochemical deposition of lithium and subsequent dendrite growth over the last several decades has been able to establish correlations between various operating conditions and dendrite microstructures (Cheng et al. 2017; Lin et al. 2017). Based on thermodynamic and kinetic understanding of the electrochemical reduction process, it is possible to develop mathematical models that can capture the experimental features observed during deposition of metallic lithium (Chen et al. 2015; Wood et al. 2016). Although nucleation and growth mechanisms may vary according to the nature of the electrode and electrolyte, the deposition mechanism of lithium metal generally can be divided into two parts:

(a) Nucleation of lithium
(b) Growth of dendritic protrusions

In this book chapter, different mathematical models used for predicting the nucleation of lithium ions will be discussed in Sect. 2. Different mathematical models to capture the growth of dendritic protrusions in contact with liquid and nonliquid (polymer and ceramic) electrolytes will be discussed in Sects. 3 and 4, respectively. Some unsolved questions that can help to better understand the metal

deposition process on top of the lithium anode will be discussed in Sect. 5. Finally, the future outlook in terms of modeling nucleation and propagation of dendrites will be summarized in Sect. 6.

## 2    Dendrite Nucleation Models

Classical nucleation theory (CNT) has been widely used to capture the lithium nucleation process on top of surfaces assumed to be free of defects (Paunovic and Schlesinger 2006). An alternative theory of instantaneous or progressive nucleation has been developed to describe the case in which surface defects provide nucleation sites for lithium dendrites (Paunovic and Schlesinger 2006; Stark et al. 2013).

Note that these theories were developed for liquid electrolyte systems. In the presence of solid electrolytes, nucleation should be affected by the extra mechanical work needed to deform the nearby electrolyte or the current collector. Under the assumption that the current collector deforms elastically, Motoyama et al. found that the overpotential for deposition increased almost linearly with Young's modulus of the current collector (Motoyama et al. 2015).

### 2.1    Classical Nucleation Theory

Under CNT, nucleation proceeds only when it decreases the total Gibbs free energy of the system (Mullin 2001). Contributions to the total free energy include electrolyte displacement, lithium deposition, and the formation of new interfaces between the surface and electrolyte and the newly deposited lithium.

Based on these considerations, Ely and García determined (Ely and García 2013) a critical radius above which a hemispherical lithium deposit becomes thermodynamically stable and below which the nucleus should spontaneously dissolve (Schmickler and Santos 2010). However, the deposition process also consumes additional energy in the form of a larger overpotential (Mullin 2001; Paunovic and Schlesinger 2006) to overcome the kinetic barrier associated with the transformation of ionic lithium (in solution) to atomic lithium (on the substrate), along with surface tension in the case of curved surfaces (Monroe and Newman 2003; Ely and García 2013). Using a modified Butler-Volmer equation incorporating the influence of surface curvature (Monroe and Newman 2003) on reaction current, Ely and García determined a kinetic critical radius at which the influences of nonnegative applied overpotentials and surface curvature are balanced. This critical radius is inversely proportional to the overpotential (Ely and García 2013) and usually larger than the critical thermodynamic radius. Stable growth is expected for lithium nuclei larger than the kinetic critical radius, although those larger than the thermodynamic critical radius can coalesce with neighboring nuclei and become stable (Ely and García 2013). Experiments have confirmed that size of lithium nuclei decreases with increasing overpotential (Sano et al. 2014; Pei et al. 2017).

## 2.2    Instantaneous and Progressive Nucleation of Lithium

When surface defects are present, lithium deposits initially on these defect sites and may grow due to further deposition of metal or through coalescence with other nearby nuclei (Paunovic and Schlesinger 2006; Stark et al. 2013). Assuming that the rate of nuclei formation on defect sites is linear in the number of available sites, the fast kinetics ("instantaneous nucleation") limit suggests that almost all preexisting active nuclei start growing simultaneously with the initiation of metal deposition (Paunovic and Schlesinger 2006; Schmickler and Santos 2010; Stark et al. 2013). The slow kinetics ("progressive nucleation") limit, on the other hand, corresponds to a gradual, approximately linear increase in the number density of nuclei over time (Stark et al. 2013). These number densities, along with assumptions about nucleus shape, can be used to estimate (Paunovic and Schlesinger 2006; Schmickler and Santos 2010; Stark et al. 2013) current density and nucleus size at earlier times where the nuclei are isolated, and later after neighboring nuclei have merged.

## 3    Dendrite Growth Models: Liquid Electrolyte

Dendritic growth of existing lithium deposits in electrochemical cells using liquid electrolyte solutions has been modeled using a variety of approaches, addressing phenomena ranging from the atomic level to that of electrochemical cells. This section surveys these techniques, beginning from the smallest length scale and proceeding through successively larger length scales.

### 3.1    Atomic-Scale Modeling

Ozhabes, Gunceler, and Arias (Ozhabes et al. 2015) constructed a density functional theory (DFT) model of crystalline surfaces in contact with electrolyte solution represented by a continuum approximation and calculated surface energies and lithium surface diffusion energy barriers for a range of SEI materials. It was found that the spatial arrangement of compounds with small anions, particularly halides, allowed for lower energy barriers and that for most compounds, larger energy barriers were associated with larger surface energies (and so greater thermodynamic stability, since dendrites increase surface area). In addition, they reported that longer times for cells to short-circuit were correlated with larger diffusion barriers, which are expected to prevent atoms from diffusing quickly enough to prevent dendrite growth.

### 3.2    Molecular Scale Modeling

Molecular-scale models are suitable for simulating situations in which small amounts of material are deposited. They have been used to explore the possibility of reducing dendrite growth rates by the use of pulsed current charging.

Mayers and coworkers created (Mayers et al. 2012) a 3D Monte Carlo (MC) simulation of dendrite growth in a hemispherical SEI region surrounding a hemispherical electrode tip. Cations diffusing within the simulation domain were allowed to deposit, with a probability determined from free energy considerations, upon collision with the electrode tip or with previously deposited metal atoms. Lower reaction probability allowed cations, on average, to diffuse more deeply through the growing coating on the electrode surface, increasing deposition density. Similarly, short charging pulses combined with long rest periods encouraged denser dendrite growth.

Aryanfar and coworkers constructed (Aryanfar et al. 2014) a similar 2D MC model of dendrite growth, but with transport of lithium ions in solution determined by a combination of diffusion and migration. Pulsed-current simulations of dendrite growth on a planar electrode suggested that pulses of appropriate length and separated by appropriate time intervals can reduce dendrite growth rates.

Although not modeling lithium deposition, Li and coworkers constructed (Li et al. 2017) a 3D molecular dynamics (MD) simulation of electrolyte solution (LiTFSI dissolved in propylene carbonate) to observe interactions among the different species. They found that the application of pulsed electric fields influenced the coordination of lithium ions by anions and that lithium ion diffusivity was improved by reduced coordination. Experimental cells that were cycled with pulsed currents showed more stable behavior and more uniform deposition relative to those cycled under constant current conditions.

## 3.3    Continuum Modeling

Simulations at the continuum scale typically involve larger, faster-growing dendrites with smooth surfaces. Models at this scale typically are formulated as partial differential equations describing the local phenomena of interest, although Yamaki et al. proposed (Yamaki et al. 1998) a qualitative lithium dendrite growth mechanism based on a wide array of experimental observations, consisting of the following steps: lithium deposits nonuniformly due to surface irregularities and nonuniform ionic conductivity of the protective surface film. The resulting mechanical stress drives transport of lithium atoms within the surface, which emerge as whiskers through breaks in the film. Finally, when the whisker coverage becomes sufficiently dense, additional lithium deposits preferentially on defects and irregularities on the whiskers.

Liu and Lu simulated (Liu and Lu 2017) the growth of a two-dimensional dendrite coated with a growing SEI layer. SEI thickness changed through electrochemical reaction and strain of the underlying surface and was reflected in resistance contributions to the surface overpotentials used to determine reaction rates. The electrode reaction was additionally influenced by a hydrostatic stress term related to the surface energy and local curvature. Relatively high reaction rates at the dendrite tip, driven by surface curvature, initially caused local SEI thinning due to expansion of the underlying lithium region. The thinned SEI region, with a lower electrical

resistance, in turn favored further reaction at the dendrite tip. At the base of the dendrite, compressive stress increased SEI thickness, which slowed local lithium deposition and created depressions relative to the surrounding material. The same mechanisms led to dendrite formation at SEI-deficient locations in otherwise flat electrode surfaces.

Applying dilute solution theory and the Poisson equation in a 1D model of a symmetric cell at steady-state, but not insisting on electroneutrality, Chazalviel observed (Chazalviel 1990) quasineutral conditions over much of the cell, but a sharp drop in potential near the cathode, with a low concentration of cations and a very low concentration of anions. Unlike in corresponding experiments, the applied potential did not affect current density as current was determined almost completely by diffusion. This discrepancy was attributed to dendrite growth, the rate of which was expected to match the anion migration speed near the cathode, based on dynamic simulations. This concept was further explored through simulations of a 2D control volume, with periodic boundary conditions, moving with the tip of a thin, straight metal dendrite growing at a steady speed. Below a critical separation distance, the solution between dendrites was depleted of ions due to dendrite growth, which proceeded at the expected rate matching the anion velocity. Above the critical separation distance, it was reasoned that the growth rate would be faster than the anion velocity, but that the presence of ions between dendrites would lead to branched growth, in turn reducing the separation distance toward the critical distance.

## 3.4    Phase Field Modeling

The dynamic models of continuum dendrite growth previously surveyed here involve calculation of precise surface positions. However, simulation is often computationally expensive for irregular interfaces (Provatas and Elder 2010; Acharya 2016). Phase field modeling can be a less expensive alternative (Chen 2002). Rather than tracking interfaces, phase field models calculate "order parameters" to represent phase identity, typically with one phase assigned a value of 0 and the other assigned a value of 1 in a two-phase system. Regions with noninteger order parameter values can be interpreted as interface regions of finite thickness. To model electrochemical systems, phase field models are typically based on a phenomenological free energy functional written in terms of field parameters (Ferguson and Bazant 2012), e.g., concentration, electric potential. For an interface-moving problem, the Cahn-Hillard equations (for variables representing conserved quantities) and the Allen-Cahn equations (for non-conserved variables) are used to express the relationship between field parameter values and the estimated total free energy. Interface movement is simulated by decreasing the total free energy of the system over time.

The application of phase field modeling to electrochemical systems is an area of ongoing research. Guyer et al. (2004a) made one of the earliest attempts to model electrochemical interface evolution by Allen-Cahn kinetics, in which they

studied the charge separation on the interface and compared to the classic Gouy-Chapman-Stern model to demonstrate effectiveness. In a subsequent paper, Guyer et al. (2004b) further explored the kinetic behavior of a phase field model for an electrodeposition process. As a typical solid-liquid interface problem, lithium dendrite growth has also been extensively modeled using the phase field method. Similarly, Cogswell (2015) investigated dendrite growth patterns by adapting the Marcus theory of microscopic charge transfer to achieve quantitative agreement with Faradaic reaction kinetics, fractal growth dimension, and radius of curvature on a zinc electrode. Cogswell also indicated that small exchange current density could reduce the growth of dendrites. To investigate electrode-electrolyte interface evolution in a lithium ion battery, Liang et al. (2012) proposed a phase field model in which they simplified the Allen-Cahn equation by assuming that the interface velocity is proportional to decay in free energy. Following the same approach, Liang et al. (Liang and Chen 2014) then studied the morphologies of lithium deposits during cell charging processes and found that the growth of lithium deposits has a very strong directional tendency, which can be greatly accelerated by the existence of a large concentration gradient. García et al. (Ely et al. 2014; Jana et al. 2015) developed a phase field model to describe the growth kinetics and distribution of lithium electrodeposits on a graphite electrode surface and found that the rate of electrodeposition at the tip of an isolated dendrite to be higher than that on the surrounding electrode. Chen et al. (2015) also proposed a new phase field model to study dendrite growth out of a surface during electrodeposition, in which they formulated Butler-Volmer kinetics for lithium deposition using the electrochemical potentials of each species. A very similar general phase field modeling approach was also developed by Bazant et al. (Ferguson and Bazant 2012; Bazant 2013; Smith and Bazant 2017), in which electrochemical processes (including generalized Butler-Volmer kinetics) as well as generalized chemical kinetics were modeled to provide a quantitative description of lithium iron phosphate phase transition behavior.

The phase field method can be used to easily model complex interface shapes and topology changes. It is particularly suited for studying phase evolution, morphology changes, or splitting and merging of diffusion fields in electrochemical systems (Chen 2002; Acharya 2016). It can combine a variety of physical effects into governing equations that can be solved by straightforward numerical methods without complicated interface handling. However, it should be recognized that the phase field method also comes with disadvantages. Although phase field models offer a deeper connection to fundamental thermodynamics than other modeling methods, applying these techniques to electrochemical systems tends to involve very stiff, high-order partial differential equations requiring small numerical time step sizes that make it impractical to simulate realistic time intervals (Provatas and Elder 2010). Multiscale numerical methods that can resolve the thin interfaces inherent in phase field models while also capturing microstructure evolution on realistic time scales are presently under development (Provatas and Elder 2010), potentially making it possible to overcome some of the present challenges of modeling electrochemical systems.

## 3.5    Cell Level Modeling

Finally, some authors have modeled dendrite growth at the cell level. Rather than modeling the growth of individual dendrites, they computed quantities such as spatially varying deposition rates.

Tang and coworkers used (Tang et al. 2009) a 2D macro-homogenous model containing a graphite negative electrode, separator, and lithium cobalt oxide positive electrode (through a boundary condition), to investigate the possibility of nonuniform lithium deposition on realistically shaped anodes. As the current density at the positive electrode interface was nonuniform, an iterative scheme was used to ensure consistency between the cell potential and the fixed applied current at each timestep. When positive and negative electrodes were of equal width, lithium deposition at the end of charging was more likely near the negative electrode edge, while increasing the negative electrode width beyond a critical distance discouraged deposition.

Wood and coworkers constructed (Wood et al. 2016) a 1D continuum-level model of a Li-Li cell. Movement of electrode/solution interfaces was determined by lithium dissolution and deposition rates, and the relationship between surface overpotentials and current densities at these interfaces was given by the Butler-Volmer equation, modified with a coefficient describing surface roughness, along with an effective rate constant representing a weighted average of rate constants for electrode surface areas with fast and slow kinetics. The surface morphology was assumed to evolve with total current passed, representing the growth and merging of neighboring lithium protrusions. A Coulombic efficiency parameter was used to account for electrical isolation of "dead lithium" during stripping. Good agreement between simulation results and corresponding cycling experiments was reported.

## 4    Impact of Mechanical Stress on Dendrite Growth: Nonliquid Electrolytes

Liquid electrolyte solutions do not provide mechanical resistance to the formation and propagation of dendrites. These can lead to formation of electrically isolated dead lithium or cause catastrophic failure by internally short circuiting the cell.

Use of nonliquid electrolytes has been proposed to reduce dendrite growth by encouraging uniform deposition on lithium metal electrodes (Monroe and Newman 2005) and eliminates the fire hazard associated with leakage of flammable liquid electrolytes (Cheng et al. 2017). From the microstructural perspective, nonliquid electrolytes can be divided into two categories: (a) structurally homogeneous and (b) structurally heterogeneous electrolytes. Most electrolytes based on poly(ethylene-oxide) (PEO) or poly (acrylonitrile) (PAN) belong to the first category, whereas solid-state electrolytes, such as the ceramic-based $Li_7La_3Zr_2O_{12}$ (LLZO) and $Li_{10}GeP_2S_{12}$ (LGPS), which possess grain-interior/grain-boundary microstructure, belong to the second category. Block-copolymers with nanoscale microstructural features belong to the first category, because these heterogeneities do not impact

the dendrite growth process (Inceoglu et al. 2014). However, micron-sized grain-interior/grain-boundary microstructural heterogeneity found in ceramic solid-state electrolytes can impact the lithium deposition process through current focusing at the electrode-electrolyte interface (Kerman et al. 2017). Hence, dendrite growth mechanisms in structurally homogeneous and heterogeneous electrolytes could differ significantly and require separate modeling approaches.

## 4.1   Structurally Homogeneous Electrolytes

Nonliquid electrolytes provide mechanical resistance in the form of compressive stress to the growth of dendritic protrusions. In addition, this compression can change their shapes from needle-like (under liquid electrolytes) to a globular (usually observed in homogeneous polymer electrolytes).

The magnitude of the compressive stress depends on the elastic modulus of the electrolyte. Increasing the electrolyte modulus by even one order of magnitude can significantly slow the growth rate of dendritic protrusions (Harry et al. 2016). Figure 1 shows that stiff block-copolymer electrolytes (SEO, consisting of PEO stiffened with polystyrene) with elastic modulus around 10 MPa at 90 °C reduce the rate of growth of lithium protrusions relative to PEO-based electrolytes, which has elastic modulus values near 1 MPa at elevated temperatures. It is clearly important to incorporate mechanical stresses in models of dendrite growth into nonliquid electrolytes.
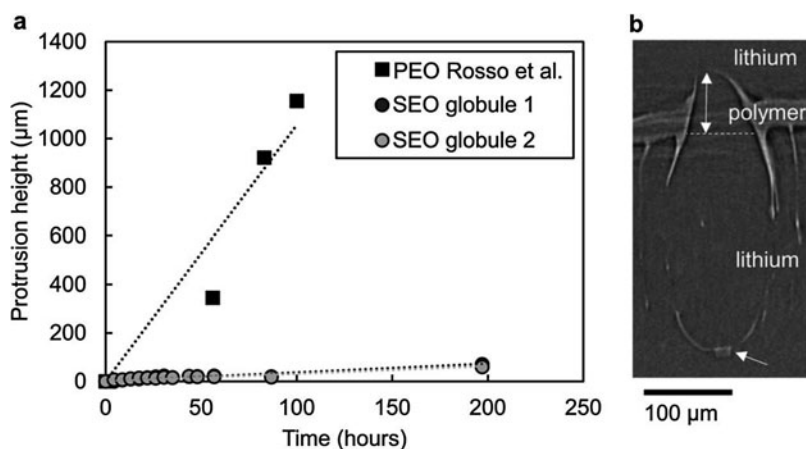


**Fig. 1** (**a**) Comparison between the rate of increase in height of dendritic protrusions observed in PEO-polymer and PEO/SEO block-copolymer-based electrolytes. Use of the SEO-based block-copolymer, with a higher elastic modulus value, resulted in significantly slower growth of the dendritic protrusion. (**b**) Dendritic protrusions were observed to be globular rather than needle-like when using high modulus polymer electrolytes. (Reproduced from Harry et al. 2016, Copyright 2016 by The Authors of Harry et al. 2016 under CC BY 4.0 terms, see http://creativecommons.org/licenses/by/4.0/)

One contribution of mechanical resistance to dendrite growth comes from electrode curvature/surface tension, which can be incorporated into the Butler-Volmer equation through its impact on the overpotential (Monroe and Newman 2003). The effect of surface curvature on dendrite growth is more prominent for polymer electrolytes than in liquids because of higher surface energy densities at lithium-polymer interfaces (about 1.5 J/m$^2$) (Monroe and Newman 2003) as compared to lithium-metal/liquid-electrolytes (about 0.1 J/m$^2$) (Liu and Lu 2017). The following Butler-Volmer equation was derived after incorporating the effects of surface curvature (Monroe and Newman 2003):

$$i_{BV} = i_{0,ref} \exp\left(\frac{2\alpha_c \gamma \overline{V}}{rRT}\right)\left[\exp\left(\frac{\alpha_a F \eta_s}{RT}\right) - \exp\left(-\frac{\alpha_c F \eta_s}{RT}\right)\right]. \qquad (1)$$

Here, $i_{BV}$ is the local reaction current density at the interface, $i_{0,ref}$ is the reference exchange current density, $\gamma$ represents surface energy density for the interface between the electrode and electrolyte, $\overline{V}$ is the partial molar volume of lithium within the electrolyte, $r$ represents the radius of the dendritic protrusion, $R$ indicates the universal gas constant, $T$ denotes temperature in Kelvin, $\eta_s$ signifies the surface overpotential, and $\alpha_a$ and $\alpha_c$ represent the anodic and cathodic transfer coefficients, respectively. The surface overpotential term is expressed as $\eta_s = \phi_s - \phi_e - U_0 + (2\gamma \overline{V}/rF)$, where $\phi_s$ is the solid phase potential, $\phi_e$ indicates the electrolyte phase potential, $U_0$ is the open circuit potential of the electrode, and $(2\gamma \overline{V}/rF)$ is the contribution of the surface curvature to the local overpotential (Monroe and Newman 2003).

The magnitude of $i_{BV}/i_{0,ref}$ as a function of the overpotential without the curvature term ($\eta = \phi_s - \phi_e - U_0$) is shown in Fig. 2. The dependence of
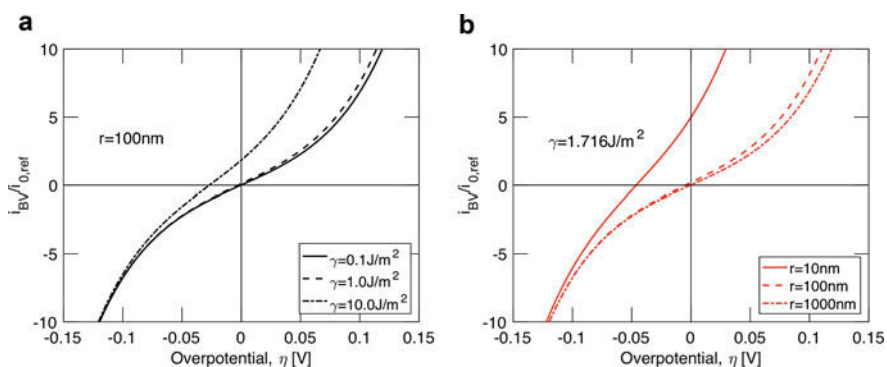


**Fig. 2** Variation in reaction current density with respect to overpotential after incorporating the effects of surface energy within the Butler-Volmer equation. (**a**) Impact of surface energy density on the overall reaction current density. Increasing surface energy enhances the anodic current. (**b**) Impact of protrusion size on the total reaction current density. Smaller lithium nuclei (r~10 nm) can easily dissolve back into the solution even under small cathodic overpotentials

the current-overpotential curve on surface energy density is shown in panel (a); increasing the value of $\gamma$ leads to higher anodic reaction current and a decreased cathodic current. Similarly, as demonstrated in panel (b), decreasing the radius of the dendritic protrusion helps to increase the anodic component of the reaction current density. Hence, the surface energy component works to stabilize the dendritic protrusions. For high surface energies ($\gamma \sim 10$ J/m$^2$), as well as extremely small dendritic protrusions (r$\sim$10 nm), anodic current is observed under small magnitudes of cathodic overpotential, which can play a significant role in the stable deposition of metals.

Furthermore, deformation of the adjacent nonliquid electrolyte was estimated to produce a local stress-induced change in electrochemical potential ($\Delta \mu_{e^-}$) (Monroe and Newman 2004, 2005):

$$\Delta \mu_{e^-}(x) = -\frac{1}{2} \left( \overline{V}_{Li} + (1 - t_{Li+}) \overline{V}_{Elec} \right) \left\{ \gamma \kappa + \vec{n} \cdot \left[ \vec{n} \cdot \left( \tau^{Li} - \tau^{Elec} \right) \right] \right\} + \\ + \frac{1}{2} \left( \overline{V}_{Li} - (1 - t_{Li+}) \overline{V}_{Elec} \right) \left( p^{Li} + p^{Elec} \right).$$

(2)

Here, $\overline{V}_{Li}$ and $\overline{V}_{Elec}$ represent the partial molar volumes of lithium and electrolyte salt, $\gamma$ is the surface energy, $\kappa$ represents local curvature, $\vec{n}$ indicates the normal vector at the lithium-electrolyte interface, $x$ denotes position on the interface, and $t_{Li+}$ stands for the lithium transference number. Also, $\tau$ represents the deviatoric stress tensor and $p$ indicates hydrostatic stress (both of which are influenced by elastic modulus values), and superscripts *Li* and *Elec* correspond to the lithium metal electrode and electrolyte, respectively. A Butler-Volmer equation incorporating the effect of this stress-induced change as well as of curvature is given in Monroe and Newman (2005). As discussed in the following subsections, several computational studies have been conducted to understand the propensity of lithium dendrite growth under the assumption of elastic or elastic-plastic deformation of the metal electrode and solid electrolyte system.

### 4.1.1 Elastic Stress Evolution in Lithium Metal Electrodes

Under the assumption of linear elastic deformation of both the metal electrode and electrolyte, Monroe and Newman solved the mechanical equilibrium equation to estimate the magnitudes of tensile and compressive stresses within the lithium metal and electrolyte, respectively (Monroe and Newman 2005). They concluded that electrolytes with shear modulus values twice that of lithium may be able to stabilize the deposition process (Monroe and Newman 2005). Following this approach, Barai et al. investigated the importance of initial stress state at the electrode-electrolyte interface (Barai et al. 2017a), introducing an alternative initial condition that could describe at least two plausible physical scenarios:

(a) A lithium protrusion begins as a small imperfection on the electrode, leading to local compressive stress upon cell assembly.
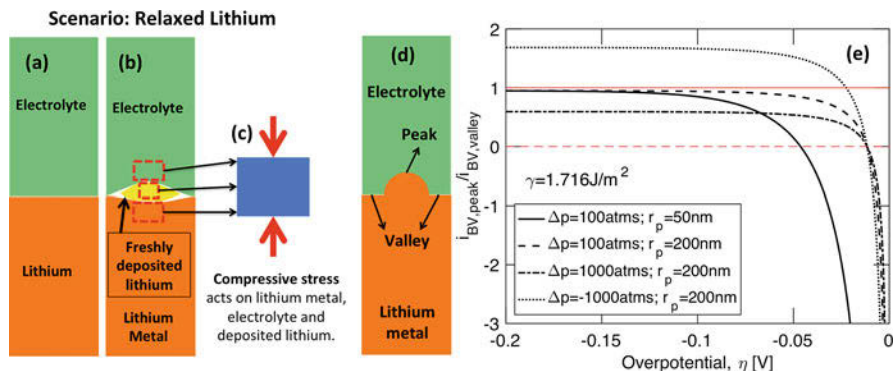
**Fig. 3** Schematic diagram demonstrating stress evolution, location of peak and valley, and ratio of current density between them. (**a**) After assembly, flat lithium metal and electrolyte in a stress-free relaxed condition. (**b**) During lithium deposition, new metal pushes the lithium downward and the electrolyte upward. (**c**) Evolution of compressive stress within electrode, electrolyte, and lithium dendrite nucleus. (**d**) Location of "peak" and "valley" for a dendritic protrusion. (**e**) Ratios of reaction current densities at the peak and valley versus reduction overpotential. Points below the solid red line correspond to stable deposition, while those above correspond to dendrite growth. The region below the dashed red line corresponds to dissolution of the peak. Hence, it may be beneficial to operate the cell at very small reduction overpotentials (around −10 mV) to eliminate preexisting dendritic protrusions. (Adopted from Barai et al. 2017a; Reproduced with permission from *Journal of the Electrochemical Society* 164 (2) A180 (2017), Copyright 2017, The Electrochemical Society)

(b) During lithium deposition, newly deposited metal pushes apart neighboring electrode and electrolyte, inducing local compressive stress (sketched in Fig. 3a–c).

Under both scenarios, lithium metal, solid electrolyte, and the dendritic protrusion experience compressive stress (Barai et al. 2017a). It should be noted that the two studies described here assumed operation at extremely low applied current densities and neglected the impact of concentration and potential gradients on the overall reaction current.

The general computational framework, developed by Monroe and Newman, assumes that the lithium nucleus appears as a small perturbation on top of the electrode surface (Monroe and Newman 2004, 2005) (see Fig. 3d). The top of the protrusion will be called the "peak," (reaction current $i_{BV, peak}$) and the flat surface surrounding the protrusion will be called the "valley" (with reaction current $i_{BV, valley}$). If the reaction current at the peak is greater than that in the valley (($i_{BV, peak}/i_{BV, valley}) > 1$), the dendritic protrusion is expected to grow. Otherwise, for ($i_{BV, peak}/i_{BV, valley}) < 1$, stable deposition of lithium is expected to occur. Ratios of total reaction current at the peak over that in the valley are shown in Fig. 3e. The solid red line indicates the stability limit. Data points lying above and below the line correspond to dendrite growth and stable deposition, respectively. In the figure, $\Delta p$ and $r_p$ indicate compressive hydrostatic stress and protrusion radius. It

is evident that increasing the magnitude of compressive stress helps to stabilize the dendritic protrusion. The size of the protrusion has a smaller impact on the current ratio at higher overpotentials. In this particular analysis, the magnitude of hydrostatic stress at the peak is double that observed in the valley. Under very low overpotentials, $\eta < 10$ mV, reaction currents at the peak and valley act in opposite directions. This can be explained by referring to Fig. 2b, which shows that at smaller reduction overpotentials, anodic reaction current occurs at the dendritic protrusion while cathodic current appears at the valley. It may be beneficial to operate the cells under these conditions to obtain very stable deposition of lithium metal.

All of the analyses previously discussed in this subsection used small-deformation mechanics. Harry et al. collected a time series of static X-ray microtomography images of a lithium dendrite growing into a solid electrolyte region and calculated mechanical stresses in the electrolyte region using large-deformation mechanics (Harry et al. 2016). Compressive stresses were found near the protrusion tip, while tensile stresses were found at the base. As shown in Fig. 3e, compressive force decreases the magnitude of reaction current density, whereas tensile stress increases it. Hence, polymer electrolyte should suppress growth at the peak and enhance growth at the base (Harry et al. 2016), consistent with formation of globular dendrites (shown in Fig. 1b).

### 4.1.2 Elastic-Plastic Stress Evolution in Lithium Metal Electrode

Under the assumption of elastic deformation, cells containing stiff electrolytes may experience very high stresses (in the range of GPa) (Monroe and Newman 2005; Barai et al. 2017a). However, stresses of this magnitude cannot be sustained by lithium metal or solid electrolytes. Ductile materials deform plastically and brittle electrolytes experience fracture beyond their yield and ultimate strengths, respectively. Hence, it is very important to model elastic-plastic deformation and rupture observed in lithium metal and electrolyte to correctly estimate stresses acting at electrode-electrolyte interfaces.

Elastic-plastic deformation has been observed in both metallic lithium and polyethylene oxide (PEO)-based electrolytes (Geng et al. 2002; Schultz 2002). Plastic deformation is irreversible in nature and consumes extra energy (see Sethuraman et al. (2010) for a similar outcome in silicon anodes). Figure 4a, b shows the experimentally observed elastic-plastic stress-strain relations that occur in lithium metal and PEO electrolyte. A nonlinear isotropic strain hardening law has been applied for both materials (Barai et al. 2017b):

$$\sigma_y = \sigma_0 + H\varepsilon_{pl}^m. \tag{3}$$

Here, $\sigma_y$ is the yield stress, $\sigma_0$ is the initial yield strength, $H$ indicates the hardening modulus, $m$ denotes the hardening exponent, and $\varepsilon_{pl}$ indicates the equivalent plastic strain. The initial yield strength, hardening modulus, and hardening exponent are material-specific parameters. Appropriate parameter values were estimated by fitting to experimental data (Barai et al. 2017b).
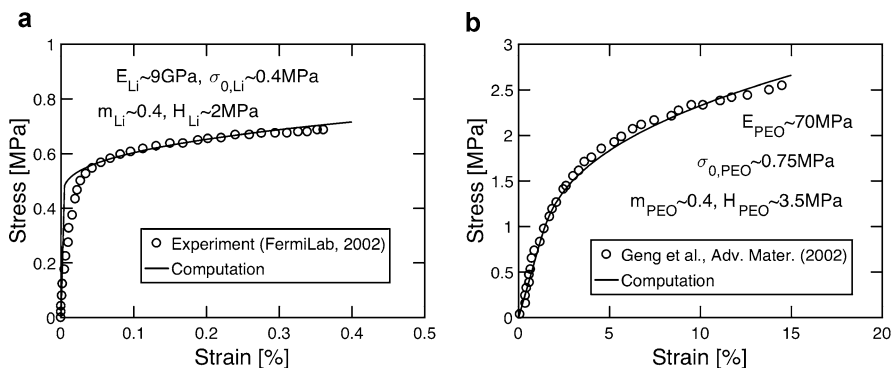
**Fig. 4** Elastic-plastic deformation observed in different materials, and the corresponding computational isotropic nonlinear hardening models and parameter values, for (**a**) lithium metal, and (**b**) PEO. (Reproduced from Barai et al. 2017b: Barai et al., *Physical Chemistry Chemical Physics* (2017) **19** 20493 with permission from the Royal Society of Chemistry)

For high-modulus polymer electrolytes, plastic deformation of the lithium dendritic protrusions and the nearby electrolyte is possible. Plastic flow of lithium from the protrusion peak to the surrounding valley region can lead to decreased protrusion height. High modulus electrolytes lead to suppression of dendritic protrusions in two different ways, as reflected in Eq. (1) (Barai et al. 2017b):

(a) The cathodic current obtained from the Butler-Volmer equation (shown by the blue line in Fig. 5a) decreases where compressive stress becomes high.
(b) Decrease in protrusion height due to plastic deformation of lithium metal reduces the component of reaction current induced by the concentration-overpotential gradient (shown by the red squares in Fig. 5a).

The combined effect is denoted by the black line in Fig. 5a, which also indicates that stable deposition of lithium may be possible if the elastic modulus of polymer is approximately 20 times larger than that of lithium.

As shown in Fig. 5b, the rapid decrease in current ratio actually occurs at lower, intermediate electrolyte modulus values, which lead to plastic deformation of lithium but elastic deformation of electrolyte. Figure 5b demonstrates that a threefold increase in the initial yield strength of present-day PEO can help significantly in suppressing growth of protrusions (Barai et al. 2017b).

Other research groups have also studied the impact of elastic plastic deformation of lithium metal on the growth of dendritic protrusions. The Butler-Volmer equation was modified by incorporating surface curvature and including a model of plastic deformation (Ferrese and Newman 2014; Jana and García 2017). The dendrite growth process was divided into two categories (Jana and García 2017): (i) tip controlled growth governed by concentration-overpotential limitations and (ii) base-driven growth that occurs due to plastic deformation of lithium. It was concluded

**Fig. 5** (**a**) With increasing electrolyte modulus, mechanical stress-induced reduction in reaction current (blue line), as well as the drop in protrusion height due to plastic deformation of lithium and corresponding impact on reaction current (red line) helps to stabilize the lithium deposition process (net effect shown by black circles). (**b**) Increasing the yield strength of present-day low modulus polymer electrolytes can effectively suppress the lithium dendritic protrusions. (Reproduced from Barai et al. 2017b: Barai et al., *Physical Chemistry Chemical Physics* (2017) **19** 20493 with permission from the Royal Society of Chemistry)
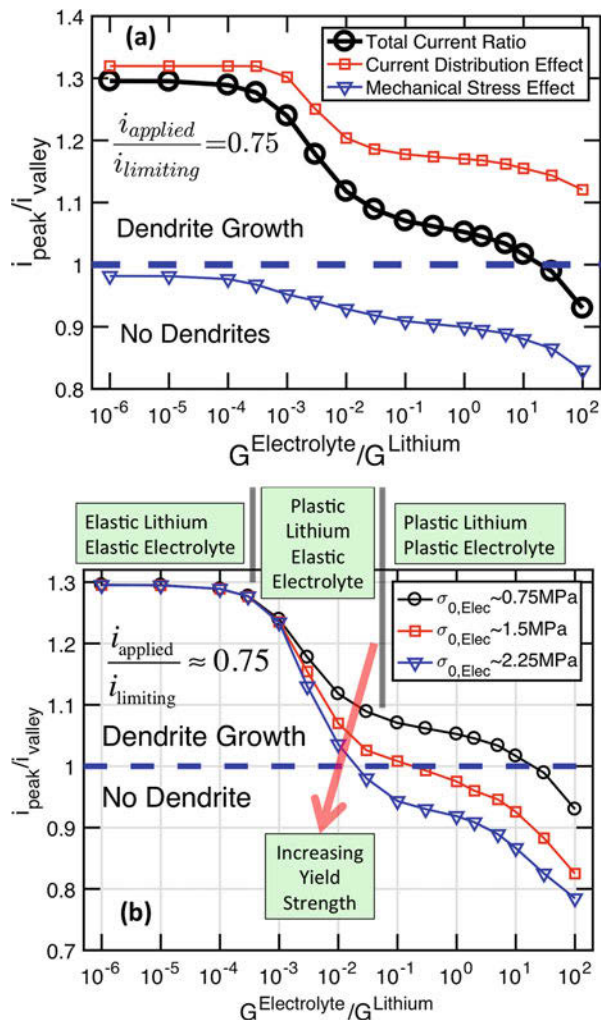
that large hydrostatic stresses and minimal deviatoric stresses are desirable for suppressing dendrite growth (Jana and García 2017). Simulations including a stiff ceramic separator (with elastic modulus in the range of tens of gigapascals) showed severe plastic deformation of metallic lithium and successful suppression of dendritic protrusions (Ferrese et al. 2012; Ferrese and Newman 2014). It was found that plastic deformation of lithium metal was much more important in suppressing the growth of dendritic protrusions than was elastic deformation or the effect of pressure on reaction kinetics (Ferrese and Newman 2014).

All three studies described here, which incorporated plastic deformation of lithium metal, assumed small-strain and linear strain-displacement relations.

However, significant dendrite growth will cause large deformations, violating the usual small-displacement assumption. Hence, it will be necessary to develop simulations that remove this constraint.

## 4.2   Structurally Inhomogeneous Electrolytes

There are other solid electrolytes that show microstructural heterogeneity. Mechanisms of lithium ion transport, plasticity, and fracture of these heterogeneous solid electrolytes can vary significantly (Stephan and Nahm 2006; Kerman et al. 2017), resulting in dendritic protrusions with different growth mechanisms and morphologies. Electrolytes with inhomogeneous microstructures can be broadly divided into two subcategories: (a) solid-state electrolytes that show distinct grain-interior/grain-boundary microstructure and (b) composite electrolytes in which a polymer matrix is reinforced with ceramic fillers. There are few computational studies on dendrite growth within these structurally inhomogeneous electrolytes, which are relatively new candidates for use in lithium ion batteries.

### 4.2.1   Single-Ion Conducting Solid State Electrolytes

In homogeneous electrolytes (both liquid and polymers), diffusion of anions helps lithium ions move through the electrolyte (Stephan and Nahm 2006). However, in ceramic electrolytes, current is carried only by the lithium ions because the anions are immobilized as part of the stable crystalline structure (Kamaya et al. 2011). Some examples of single ion conducting electrolyte are LLZO and LGPS, which are oxide-based and sulfide-based ceramics, respectively. However, for chemical stability, they are often doped with other elements, such as gallium or aluminum (Rettenwander et al. 2016). Ceramic solid-state electrolytes have extremely high elastic modulus values, making them suitable for preventing growth of dendritic protrusions; LGPS and LLZO have elastic modulus values of around 20 GPa and 150 GPa (Kerman et al. 2017), respectively. However, all of the solid-state electrolytes rupture in a brittle fashion under sufficient stress.

The impact of mechanical stress on dendrite growth was derived earlier under the assumption of homogeneous electrolyte (Monroe and Newman 2004). This model was modified to investigate the impact of dendrite growth within ceramic electrolytes (Ahmad and Viswanathan 2017). Significantly decreased partial molar volume of lithium ions within solid-state-electrolytes impacts the dendrite growth process. It was concluded that dendritic protrusions should not grow for electrolytes softer than lithium (Ahmad and Viswanathan 2017). Due to anisotropy in crystal structure, mechanical stiffness tensors are typically anisotropic. It was further concluded that the anisotropic behavior does not affect the overall regime of stability, but rather simply shifts the stability limits (Ahmed and Viswanathan 2017).

An equilibrium-based computational model demonstrated that propagation of slender dendritic protrusions rupturing a path through the solid-state electrolyte is energetically favorable relative to deposition of lithium on the other side of the electrolyte (Porz et al. 2017), which has also been supported by experimental

observations. The brittle rupture phenomena and low magnitudes of fracture energy were suggested as major factors contributing to the easy growth of lithium dendrites within ceramic single ion conductors. Also, the lack of flexibility of the solid-state electrolytes led to delamination at the electrode-electrolyte interface, which can result in high overpotentials.

### 4.2.2 Composite Polymer-Ceramic Electrolytes

The standard polymer electrolytes (such as PEO) possess neither sufficiently high conductivity nor elastic modulus values for sustained operation with lithium metal anodes at room temperature. However, their amorphous regions conduct lithium ions more readily than do their crystalline regions, so elevating the operating temperature above the melting temperature improves lithium ion conduction. Alternatively, ceramic fillers (such as micron-sized or nanometer-sized $Al_2O_3$ or $SiO_2$) are sometimes introduced to produce polymer-ceramic composites with higher conductivity and elastic modulus at room temperature, which inhibits the crystallization of polymer chains (Stephan and Nahm 2006).

Some nanometer-sized filler particles, such as those composed of $SiO_2$, form three-dimensional networks in which anions are immobilized (Schaefer et al. 2013), increasing the transference number of the electrolyte. Tikekar and coworkers constructed a steady-state model of a cell containing an electrolyte region for which mechanical stress was neglected, but with a specified fraction of fixed anions (Tikekar et al. 2014). During deposition at high current densities, they observed steep potential changes and ion depletion near the cathode, which is thought to drive dendrite growth. Increasing the fraction of stationary anions decreased the electric field strength in this region, particularly at high current densities, which was expected to promote stable electrode surfaces.

## 5   Questions

Even though a significant amount of modeling work has been conducted in the past several decades to understand the mechanisms behind nucleation and growth of lithium dendrites, present understanding is far from satisfactory. The fact that no material or strategy has been developed that can completely stabilize the lithium deposition process reflects this remarkable gap in knowledge. Scientists have been able to develop continuum models that can explain the voltage and current profiles observed during the deposition of lithium (Barton and Bockris 1962; Diggle et al. 1969; Motoyama et al. 2015; Wood et al. 2016; Pei et al. 2017). Other computational models based on phase field and Monte Carlo techniques have also been able to predict the evolution of dendrite morphology under different operating conditions (Chen et al. 2015). Unfortunately, these have not led to stable deposition of lithium on metal electrodes. It is important to ask the correct questions to make reasonable technological progress. A few such questions that scientists should focus on for taking the lithium metal anode technology to its next level are provided here.

It is important to recognize that successful commercial implementation of lithium metal anodes requires stable deposition for thousands of charge-discharge cycles. Dendrites may not grow sufficiently large to short the cell during a single charge step. Usually, multiple charge-discharge cycles are needed for a dendritic protrusion to puncture through the separator and short the cell (Stone et al. 2012; Sharafi et al. 2016). Hence, stable deposition for a single charge-discharge cycle may not guarantee dendrite-free operation. However, a majority of the computational studies attempt to model the growth of dendrites during a single deposition process (Monroe and Newman 2003; Ely et al. 2014). Even if it is possible to stabilize the deposition of lithium on top of flat surfaces, nonuniform dissolution of lithium during discharge may lead to development of sites for dendrite growth during the subsequent charge steps (Wood et al. 2016). Focus should be concentrated on not only the first deposition of lithium, but also the later dissolution and plating steps.

A solid-electrolyte-interphase (SEI) develops at the lithium-electrolyte interface as soon as the lithium metal comes in contact with the organic electrolyte. SEI layer compositions are not known exactly (Cheng et al. 2016). The presence of different domains, adjacent to a metal electrode, with variable lithium ion diffusivity and conductivity can lead to nonuniform deposition at the lithium-SEI interface. Arguments have been made that rupture of the SEI layer allows for enhanced deposition of lithium and dendritic protrusions initiate from these regions (Aurbach 2000). On the other hand, it has also been argued that presence of mechanically stiff lithium-fluoride within the SEI layer can stabilize the deposition process for tens of cycles. Hence, the impact of the metal-electrolyte interphase on lithium deposition has not been completely elucidated. How the SEI layer alters the lithium dissolution process is also not clearly understood. Questions exist regarding how the lithium ions and solvent molecules migrate and diffuse through the SEI layer before reacting with the lithium electrode. Proper computational models should be developed to reconcile the effect of the SEI layer on the lithium dendrite growth mechanism.

From an experimental perspective, it has been observed that moss-like structures form during deposition of lithium at small current densities, whereas needle-shaped dendrites evolve during operation at large current (Bai et al. 2016). Dendritic microstructures obtained through computational means do not always correlate well with the experimental observations. For example, phase-field modeling techniques produce linear dendritic structures at low overpotential, and tree-like structures at higher overpotentials (Chen et al. 2015). This seems not to correlate well with experimental observations. Also, other computational methodologies have observed formation of dead lithium even under cathodic overpotentials (Jana et al. 2015), which is very unlikely in realistic situations. Hence, future models should focus on predicting scenarios that correlate more closely with the experimental observations. Another important feature that has not been thoroughly investigated is the crystal orientation of the lithium surface on which deposition occurs. Proper first-principles-based analysis should be conducted to determine the surface energies of different lithium crystal orientations. The propensity of dendrite formation during deposition on different types of lithium crystal surfaces should be appropriately elucidated.

Limited room temperature conductivity has been a major issue preventing widespread implementation of nonliquid electrolytes in commercial devices (Stephan and Nahm 2006). Recent developments have yielded ceramic solid state electrolytes with room temperature conductivity even larger than that of liquids (Kato et al. 2016). However, there are no computational models that analyze the growth of lithium dendrites within ceramic electrolytes. Some phenomenological models have been developed to elucidate the higher propensity of lithium dendrite growth within ceramic electrolytes (Porz et al. 2017). Dissolution of these dendritic protrusions under anodic overpotential should also be thoroughly investigated. Significant questions remain regarding the cycle performance of lithium anodes with ceramic electrolytes, for example, on the impact of grain boundaries and surface flaws on the overall lithium deposition-dissolution process (Kerman et al. 2017). The abilities of polymer-based solid electrolytes to stabilize the electrochemical deposition of lithium should also be investigated further through computational models. Successful dendrite prevention mechanisms can come from computational models if the physical phenomena of nucleation and growth are understood.

# 6    Outlook

Given the advantages of lithium metal anodes, associated with their low electrochemical potential and high gravimetric capacity, next-generation lithium ion batteries should aim to use these anode materials to increase energy density. The largest issues with lithium metal show up in the form of dendrite growth, formation of dead lithium, and continuous decomposition of electrolyte at the solid-electrolyte interphase layer. Computational approaches should take a leading role in devising strategies to solve these problems. Since the success of a computational model depends on the detailed capabilities of the adopted theory, a good understanding of the physico-chemical phenomena is of utmost importance for the development of an accurate theoretical framework. Taking into consideration that several physical phenomena can occur at the lithium-electrolyte interface, such as mass transport, charge balance, heat transfer, and mechanical equilibrium, multiphysics computational models should be developed to successfully predict the lithium dendrite growth process. Furthermore, the physical phenomena observed during growth of a dendritic protrusion span multiple length scales. For example, nucleation of lithium occurs at the nanometer level. The cathode and the anode are usually placed several microns apart in a cell. To cause a short circuit, a dendrite likewise has to grow by several microns to reach the cathode. To capture all the phenomena that occur over multiple length scales, multiscale computational models should be developed.

DFT models have been widely developed and utilized to understand the behavior of lithium and electrolyte separately, as well as at their interfaces. However, these models are capable of predicting only local features at an atomic scale. The number of atoms required to form a lithium dendrite nucleus is also excessively large for prediction using the DFT and ab-initio molecular dynamic (AIMD) simulation techniques. Hence, it is necessary to use either classical molecular dynamics

(CMD) or Monte Carlo (MC)-based computational procedures for understanding the physical phenomena occurring at nano-scale, with properties measured using the AIMD procedures.

On the other hand, well-developed computational techniques exist for modeling the various physical phenomena observed at continuum (or micron sized) length scales. Usually these continuum models have parameters that need to be determined from experiments or lower length scale simulations.

The majority of bulk materials appear to be homogeneous at the continuum level. However, there may be heterogeneities at the micron or submicron length scales, also known as the mesoscale, which are not visible in the continuum level. The DFT/AIMD-based lower length scale simulations are also not affected by these heterogeneities due to their localized nature and provide parameters for perfect crystals. Realistically, these microstructural heterogeneities give rise to higher stress concentrations or current focusing, which leads to the initiation of dendrites and subsequent failure of cells. Successfully bridging these computational techniques applicable at different length scales, through modeling of mesoscale phenomena, is of utmost importance for developing accurate models. Appropriate predictions of features observed at the mesoscale can lift the significance of computational models to the next level. Faster supercomputers with high computational power can also help to bridge the gap between atomistic and continuum length scales through the modeling of physical phenomena observed at the mesoscale.

# References

Acharya N (2016) Phase field modeling of electrodeposition process in lithium metal batteries. Master of Science, Missouri University of Science and Technology

Ahmad Z, Viswanathan V (2017) Stability of electrodeposition at solid-solid interfaces and implications for metal anodes. Phys Rev Lett 119(5):056003

Ahmed Z, Viswanathan V (2017) Role of anisotropy in determining stability of electrodeposition at solid-solid interfaces. Phys Rev Mater 1:055403

Aryanfar A, Brooks D, Merinov BV, Goddard WA, Colussi AJ, Hoffmann MR (2014) Dynamics of lithium dendrite growth and inhibition: pulse charging experiments and Monte Carlo calculations. J Phys Chem Lett 5(10):1721–1726

Aurbach D (2000) Review of selected electrode-solution interactions which determine the performance of Li and Li ion batteries. J Power Sources 89(2):206–218

Bai P, Li J, Brushett FR, Bazant MZ (2016) Transition of lithium growth mechanisms in liquid electrolytes. Energy Environ Sci 9(10):3221–3229

Barai P, Higa K, Srinivasan V (2017a) Effect of initial state of lithium on the propensity for dendrite formation: a theoretical study. J Electrochem Soc 164(2):A180–A189

Barai P, Higa K, Srinivasan V (2017b) Lithium dendrite growth mechanisms in polymer electrolytes and prevention strategies. Phys Chem Chem Phys 19(31):20493–20505

Barton JL, Bockris JOM (1962) The electrolytic growth of dendrites from ionic solutions. Proc R Soc A: Math Phys Eng Sci 268(A):485–505

Bazant MZ (2013) Theory of chemical kinetics and charge transfer based on nonequilibrium thermodynamics. Acc Chem Res 46(5):1144–1160

Blomgren GE (2017) The development and future of lithium ion batteries. J Electrochem Soc 164(1):A5019–A5025

Chazalviel JN (1990) Electrochemical aspects of the generation of ramified metallic electrodeposits. Phys Rev A 42(12):7355–7367

Chen LQ (2002) Phase-field models for microstructure evolution. Annu Rev Mater Res 32:113–140

Chen L, Zhang HW, Liang LY, Liu Z, Qi Y, Lu P, Chen J, Chen LQ (2015) Modulation of dendritic patterns during electrodeposition: a nonlinear phase-field model. J Power Sources 300:376–385

Cheng XB, Zhang R, Zhao CZ, Wei F, Zhang JG, Zhang Q (2016) A review of solid electrolyte interphases on lithium metal anode. Adv Sci 3(3):1500213

Cheng XB, Zhang R, Zhao CZ, Zhang Q (2017) Toward safe lithium metal anode in rechargeable batteries: a review. Chem Rev 117(15):10403–10473

Cogswell DA (2015) Quantitative phase-field modeling of dendritic electrodeposition. Phys Rev E 92(1):011301

Diggle JW, Despic AR, Bockris JOM (1969) The mechanism of the dendritic electrocrystallization of zinc. J Enectrochem Soc 116(11):1503–1514

Ely DR, García RE (2013) Heterogeneous nucleation and growth of lithium electrodeposits on negative electrodes. J Electrochem Soc 160(4):A662–A668

Ely DR, Jana A, García RE (2014) Phase field kinetics of lithium electrodeposits. J Power Sources 272:581–594

Ferguson TR, Bazant MZ (2012) Nonequilibrium thermodynamics of porous electrodes. J Electrochem Soc 159(12):A1967–A1985

Ferrese A, Newman J (2014) Mechanical deformation of a lithium-metal anode due to a very stiff separator. J Electrochem Soc 161(9):A1350–A1359

Ferrese A, Albertus P, Christensen J, Newman J (2012) Lithium redistribution in lithium-metal batteries. J Electrochem Soc 159(10):A1615–A1623

Gallagher KG, Goebel S, Greszler T, Mathias M, Oelerich W, Eroglu D, Srinivasan V (2014) Quantifying the promise of lithium-air batteries for electric vehicles. Energy Environ Sci 7(5):1555–1563

Geng HZ, Rosen R, Zheng B, Shimoda H, Fleming L, Liu J, Zhou O (2002) Fabrication and properties of composites of poly(ethylene oxide) and functionalized carbon nanotubes. Adv Mater 14(19):1387–1390

Guyer JE, Boettinger WJ, Warren JA, McFadden GB (2004a) Phase field modeling of electrochemistry. I. Equilibrium. Phys Rev E 69(2):021603

Guyer JE, Boettinger WJ, Warren JA, McFadden GB (2004b) Phase field modeling of electrochemistry. II. Kinetics. Phys Rev E 69(2):021604

Harry KJ, Higa K, Srinivasan V, Balsara NP (2016) Influence of electrolyte modulus on the local current density at a dendrite tip on a lithium metal electrode. J Electrochem Soc 163(10):A2216–A2224

Inceoglu S, Rojas AA, Devaux D, Chen XC, Stone GM, Balsara NP (2014) Morphology-conductivity relationship of single-ion-conducting block copolymer electrolytes for lithium batteries. ACS Macro Lett 3(6):510–514

Jana A, García RE (2017) Lithium dendrite growth mechanisms in liquid electrolytes. Nano Energy 41:552–565

Jana A, Ely DR, García RE (2015) Dendrite-separator interactions in lithium-based batteries. J Power Sources 275:912–921

Kamaya N, Homma K, Yamakawa Y, Hirayama M, Kanno R, Yonemura M, Kamiyama T, Kato Y, Hama S, Kawamoto K, Mitsui A (2011) A lithium superionic conductor. Nat Mater 10(9):682–686

Kato Y, Hori S, Saito T, Suzuki K, Hirayama M, Mitsui A, Yonemura M, Iba H, Kanno R (2016) High-power all-solid-state batteries using sulfide superionic conductors. Nat Energy 1:16030

Kerman K, Luntz A, Viswanathan V, Chiang YM, Chen ZB (2017) Review-practical challenges hindering the development of solid state Li ion batteries. J Electrochem Soc 164(7):A1731–A1744

Khurana R, Schaefer JL, Archer LA, Coates GW (2014) Suppression of lithium dendrite growth using cross-linked polyethylene/poly(ethylene oxide) electrolytes: a new approach for practical lithium-metal polymer batteries. J Am Chem Soc 136(20):7395–7402

Li Z, Huang J, Liaw BY, Metzler V, Zhang JB (2014) A review of lithium deposition in lithium-ion and lithium metal secondary batteries. J Power Sources 254:168–182

Li Q, Tan S, Li LL, Lu YY, He Y (2017) Understanding the molecular mechanism of pulse current charging for stable lithium-metal batteries. Sci Adv 3(7):e1701246

Liang LY, Chen LQ (2014) Nonlinear phase field model for electrodeposition in electrochemical systems. Appl Phys Lett 105(26):263903

Liang LY, Qi Y, Xue F, Bhattacharya S, Harris SJ, Chen LQ (2012) Nonlinear phase-field model for electrode-electrolyte interface evolution. Phys Rev E 86(5):051609

Lin DC, Liu YY, Cui Y (2017) Reviving the lithium metal anode for high-energy batteries. Nat Nanotechnol 12(3):194–206

Liu GY, Lu W (2017) A model of concurrent lithium dendrite growth, SEI growth, SEI penetration and regrowth. J Electrochem Soc 164(9):A1826–A1833

Lu YY, Das SK, Moganty SS, Archer LA (2012) Ionic liquid-nanoparticle hybrid electrolytes and their application in secondary lithium-metal batteries. Adv Mater 24(32):4430–4435

Mayers MZ, Kaminski JW, Miller TF (2012) Suppression of dendrite formation via pulse charging in rechargeable lithium metal batteries. J Phys Chem C 116(50):26214–26221

Monroe C, Newman J (2003) Dendrite growth in lithium/polymer systems – a propagation model for liquid electrolytes under galvanostatic conditions. J Electrochem Soc 150(10):A1377–A1384

Monroe C, Newman J (2004) The effect of interfacial deformation on electrodeposition kinetics. J Electrochem Soc 151(6):A880–A886

Monroe C, Newman J (2005) The impact of elastic deformation on deposition kinetics at lithium/polymer interfaces. J Electrochem Soc 152(2):A396–A404

Motoyama M, Ejiri M, Iriyama Y (2015) Modeling the nucleation and growth of li at metal current collector/LiPON interfaces. J Electrochem Soc 162(13):A7067–A7071

Mullin JW (2001) Crystallization. Butterworth Heinemann, Boston

Myung ST, Maglia F, Park KJ, Yoon CS, Lamp P, Kim SJ, Sun YK (2017) Nickel-rich layered cathode materials for automotive lithium-ion batteries: achievements and perspectives. ACS Energy Lett 2(1):196–223

Nitta N, Yushin G (2014) High-capacity anode materials for lithium- ion batteries: choice of elements and structures for active particles. Part Part Syst Charact 31(3):317–336

Nitta N, Wu FX, Lee JT, Yushin G (2015) Li-ion battery materials: present and future. Mater Today 18(5):252–264

Ozhabes Y, Gunceler D, and Arias TA (2015) Stability and surface diffusion at lithium-electrolyte interphases with connections to dendrite suppression. arXiv:1504.05799 [cond-mat.mtrl-sci]

Paunovic M, Schlesinger M (2006) Fundamentals of electrochemical deposition. Wiley Inter-science, Wiley

Pei A, Zheng GY, Shi FF, Li YZ, Cui Y (2017) Nanoscale nucleation and growth of electrode-posited lithium metal. Nano Lett 17(2):1132–1139

Peng Z, Wang SW, Zhou JJ, Jin Y, Liu Y, Qin YP, Shen C, Han WQ, Wang DY (2016) Volumetric variation confinement: surface protective structure for high cyclic stability of lithium metal electrodes. J Mater Chem A 4(7):2427–2432

Porz L, Swamy T, Sheldon BW, Rettenwander D, Fromling T, Thaman HL, Berendts S, Uecker R, Carter WC, Chiang YM (2017) Mechanism of lithium metal penetration through inorganic solid electrolytes. Adv Energy Mater 7(20):1701003

Provatas N, Elder K (2010) Phase field methods in material science and engineering. Wiley-VCH, Weinheim

Rahn CD (2013) Battery systems engineering. Wiley, Chichester

Rettenwander D, Redhammer G, Preishuber-Pflugl F, Cheng L, Miara L, Wagner R, Welzl A, Suard E, Doeff MM, Wilkening M, Fleig J, Amthauer G (2016) Structural and electrochemical consequences of Al and Ga Cosubstitution in $Li_7La_3Zr_2O_{12}$ solid electrolytes. Chem Mater 28(7):2384–2392

Sano H, Sakaebe H, Senoh H, Matsumoto H (2014) Effect of current density on morphology of lithium electrodeposited in ionic liquid-based electrolytes. J Electrochem Soc 161(9):A1236–A1240

Schaefer JL, Yanga DA, Archer LA (2013) High lithium transference number electrolytes via creation of 3-dimensional, charged, Nanoporous networks from dense functionalized nanoparticle composites. Chem Mater 25(6):834–839

Schmickler W, Santos E (2010) Metal deposition and dissolution. Springer, Berlin

Schultz, R (2002) Lithium: measurement of young's modulus and yield strength. Fermilab-TM-2191:1–6

Sethuraman VA, Srinivasan V, Bower AF, Guduru PR (2010) In situ measurements of stress-potential coupling in lithiated silicon. J Electrochem Soc 157(11):A1253–A1261

Sharafi A, Meyer HM, Nanda J, Wolfenstine J, Sakamoto J (2016) Characterizing the Li-Li7La3Zr2O12 interface stability and kinetics as a function of temperature and current density. J Power Sources 302:135–139

Shin WK, Kannan AG, Kim DW (2015) Effective suppression of dendritic lithium growth using an ultrathin coating of nitrogen and sulfur Codoped graphene Nanosheets on polymer separator for lithium metal batteries. ACS Appl Mater Interfaces 7(42):23700–23707

Smith RB, Bazant MZ (2017) Multiphase porous electrode theory. J Electrochem Soc 164(11):E3291–E3310

Stark JK, Ding Y, Kohl PA (2013) Nucleation of electrodeposited lithium metal: dendritic growth and the effect of co-deposited sodium. J Electrochem Soc 160(9):D337–D342

Stephan AM, Nahm KS (2006) Review on composite polymer electrolytes for lithium batteries. Polymer 47(16):5952–5964

Stone GM, Mullin SA, Teran AA, Hallinan DT, Minor AM, Hexemer A, Balsara NP (2012) Resolution of the modulus versus adhesion dilemma in solid polymer electrolytes for rechargeable lithium metal batteries. J Electrochem Soc 159(3):A222–A227

Takehara Z (1997) Future prospects of the lithium metal anode. J Power Sources 68(1):82–86

Tang M, Albertus P, Newman J (2009) Two-dimensional modeling of lithium deposition during cell charging. J Electrochem Soc 156(5):A390–A399

Tikekar MD, Archer LA, Koch DL (2014) Stability analysis of electrodeposition across a structured electrolyte with immobilized anions. J Electrochem Soc 161(6):A847–A855

Wood KN, Kazyak E, Chadwick AF, Chen KH, Zhang JG, Thornton K, Dasgupta NP (2016) Dendrites and pits: untangling the complex behavior of lithium metal anodes through operando video microscopy. Acs Cent Sci 2(11):790–801

Xu W, Wang JL, Ding F, Chen XL, Nasybutin E, Zhang YH, Zhang JG (2014) Lithium metal anodes for rechargeable batteries. Energy Environ Sci 7(2):513–537

Yamaki J, Tobishima S, Hayashi K, Saito K, Nemoto Y, Arakawa M (1998) A consideration of the morphology of electrochemically deposited lithium in an organic electrolyte. J Power Sources 74(2):219–227

Zheng GY, Lee SW, Liang Z, Lee HW, Yan K, Yao HB, Wang HT, Li WY, Chu S, Cui Y (2014) Interconnected hollow carbon nanospheres for stable lithium metal anodes. Nat Nanotechnol 9(8):618–623

# Part VIII
# Stochastic, Coarse-Grained Models of Materials Mechanics

# Incorporating the Element of Stochasticity in Coarse-Grained Modeling of Materials Mechanics

# 55

Eric R. Homer, Ying Chen, and Christopher A. Schuh

## Contents

**Abstract**

Materials are, by their very nature, stochastic. Modeling materials across scales requires models that capture this inherent stochasticity. In this chapter, preceding a section on stochastic, coarse-grained models, we examine the elements of stochasticity and coarse-graining and the different implementations of each. Examples of the methods are also briefly discussed.

E. R. Homer (✉)
Department of Mechanical Engineering, Brigham Young University, Provo, UT, USA
e-mail: eric.homer@byu.edu

Y. Chen
Department of Materials Science and Engineering, Rensselaer Polytechnic Institute,
Troy, NY, USA
e-mail: cheny20@rpi.edu

C. A. Schuh
Department of Materials Science and Engineering, Massachusetts Institute of Technology,
Cambridge, MA, USA
e-mail: schuh@mit.edu

# 1    Introduction

In a letter to Max Born, Albert Einstein is famously quoted as saying, "Quantum mechanics is certainly imposing ... but does not really bring us any closer to the secret of the 'old one'. I, at any rate, am convinced that *He* is not playing at dice." (Quote obtained from English translation of Born-Einstein letters (1971).) Einstein was dissatisfied with the probabilistic nature of quantum mechanics and spent most of his later years searching for a unified field theory that could provide a deterministic explanation of matter. In spite of his efforts, overwhelming evidence has been presented in favor of the probabilistic mechanics that govern atomic processes. It is clear also that the stochasticity inherent in these atomic processes influences material processes at larger scales.

Consider the structures and processes presented by Tadmor and Miller (2011) in Fig. 1. These structures and processes span numerous time- and length-scales,
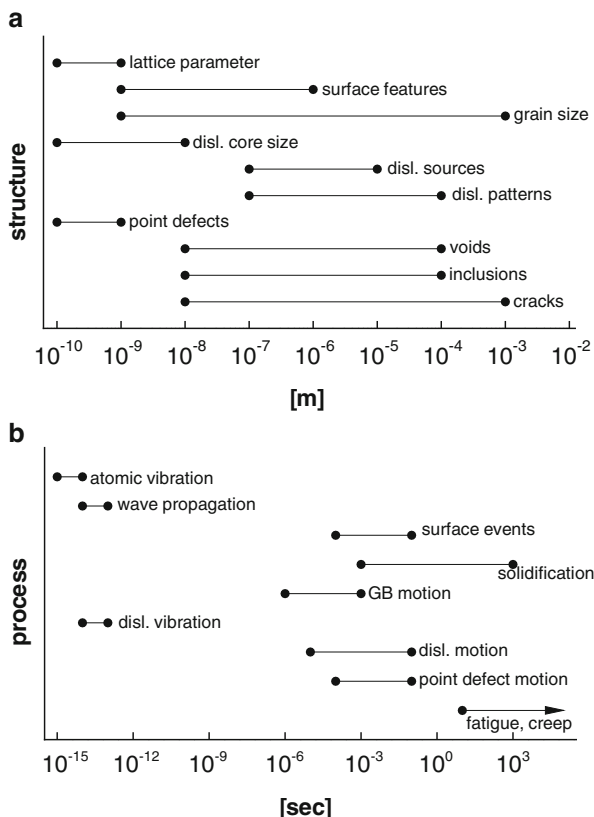


**Fig. 1** (**a**) Length-scales and (**b**) time-scales associated with different structures and processes. (Used with Permission from Tadmor and Miller 2011)

which lead to significant challenges when attempting to model specific materials or processes. Furthermore, the inherent stochasticity of the processes leads to significant challenges in resolving the underlying causes of the behavior. Consider the following behaviors that are influenced by the stochasticity of their underlying collective process:

- Fatigue crack growth: While many of the influencing factors surrounding fatigue cracking are known, crack nucleation and propagation are collective processes that have significant randomness and variability that complicate accurate modeling and prediction (Kozin and Bogdanoff 1989; Ellyin 1997; Beden et al. 2009).
- Strength at small scales: The variability of atomic structure at small scales plays a significant role in the strength observed (Uchic et al. 2004; Packard et al. 2010; Askari et al. 2015). Different microstructural features clearly play a role in this variability, but quantifying the exact influence of each of these features remains a challenge.
- Martensitic transformations: Martensitic transformations in steels and shape-memory alloys are complex, collective processes whose crystallography and influencing factors are fairly well understood. Nevertheless, prediction of which regions transform, and to a lesser extent the variant to which they transform, remains a challenge (Christian 2002; Bhattacharya 2012; Miyamoto et al. 2013).

The above and many more similar phenomena require models that can bridge the so-called meso-scales between atomic-level structure and processes and macroscopic behavior. These models must account for the increased complexity and variability of material behaviors in this range (Yip and Short 2013; Provis 2015) and represent an area of unprecedented challenge and opportunity (Hemminger et al. 2012). This handbook contains three other sections that are focused on bridging this meso-scale divide. These sections include (i) Crystal Plasticity: Atomistics to Macroscale (Cai and Ghosh), (ii) Long Timescale Atomistic Simulations: Accelerated Molecular Dynamics and Adaptive Kinetic Monte Carlo (Perez and Uberuaga), and (iii) Modeling of Microstructure Evolution: Mesoscale Challenges (Stan and Sarrao). Like the techniques presented in this section, these sections also include unique and clever solutions to address the meso-scale challenge.

In this section, we discuss meso-scale modeling approaches that coarse-grain material phenomena. But, in contrast to many meso-scale approaches, we emphasize the inclusion of stochasticity and its critical role in modeling different processes that inherit from the probabilistic nature of matter.

The coarse-graining has the advantage of representing the effective behavior of complex structures in a simpler way; individual atoms and their bonding do not need to be explicitly considered. It also allows modeling of behavior at scales comparable to those employed in experiments.

The stochasticity is incorporated as a result of the fact that so many collective processes in materials appear to behave in a random or probabilistic fashion. Furthermore, many of the processes of interest are infrequent from the perspective

of atomic vibrations. As such, our current understanding and models struggle to explain these processes using deterministic approaches.

In this overview chapter, we detail the coarse-graining and stochastic elements that characterize this class of models that are able to simulate collective phenomena at larger length-scales and on longer time-scales while still accounting for the stochastic nature of the underlying atomic processes. These two elements can be implemented in a variety of ways, and as such, each element is discussed individually. This is followed by a brief summary of implementations of these methods in the chapters included in this section of the handbook, as well as examples available in the literature. Finally, we close with an outlook on stochastic, coarse-grained modeling and discuss challenges and opportunities that exist.

## 2 Elements of Stochastic, Coarse-Grained Models

In defining the uniqueness of these meso-scale models, we find that two components set these models apart. These two components are stochasticity and coarse-graining. In the following sections, we individually discuss the different ways these two elements are often implemented in models.

## 2.1 Coarse-Graining

In all multi-scale modeling efforts, phenomena have to be modeled at different scales in a manner that captures the appropriate behaviors. Coarse-graining is one efficient way of modeling materials at larger length-scales and on longer time-scales. In coarse-graining, atomic behaviors do not have to be considered; instead, their local collective behavior is properly represented in the model with a reduced number of degrees of freedom.

We note the following ways that models frequently implement coarse-graining:

1. Coarse-graining specific objects
2. Coarse-graining regions of material

### 2.1.1 Coarse-Graining Specific Objects

Coarse-graining is often implemented because certain material phenomena are associated with specific structures. For example, many material processes involve material defects, such as vacancies, solute atoms, dislocations, interfaces, etc. In these cases, the behavior of the material is governed by the collective behavior of the atoms in the defect. For example, one can model a dislocation as a line defect rather than the collection of atoms that make up that defect. As such, one can more efficiently model the collective behavior of the defect as an object rather than modeling the individual atomic motions that make up the defect. However, in order to properly capture the physics of the coarse-grained object, it is often necessary to discretize the object on the coarser scale. For example, while treating dislocations

as line defects, dislocation dynamics models usually discretize dislocation lines by nodes or segments; each node or segment can be viewed as representing the collective behavior of atoms making up that segment.

### 2.1.2    Coarse-Graining Regions of Material

In some cases, the material phenomena may not be tied to a specific structure, or the structure may need to be subdivided. For example, in amorphous materials, it is difficult to identify a unit structure for coarse graining. Or in larger defects like grain boundaries or inclusions, one may wish to subdivide the defect into smaller regions. In this case, one can simply coarse-grain a region of material.

When coarse-graining a region of material, one must understand the implications of the length-scale selected. The challenge here is that certain processes that one may wish to model will occur over different length-scales. For example, many processes involve incubation, nucleation, propagation, and growth stages. The nucleation stage can be small while the propagation and/or growth stages involve large regions of material. Coarse-graining can involve multiple scales, but it is usually convenient to pick a single length-scale in order to keep the model simple. Again, the coarse-grained representation of the material may be discretized in the model, and the discretization scheme could change during the simulation to better capture the evolution in material property and in heterogeneous fields. In selecting a region to model any of these processes, one must use care and ensure that the underlying material process is satisfactorily captured; one must determine the ideal balance between resolution and efficiency in modeling.

## 2.2    Stochasticity

As noted in the introduction, many material processes at the meso-scale have collective processes that occur in a random or probabilistic manner. The randomness could be attributed to our incomplete understanding of the physics governing these processes, thereby causing a deterministic process to appear stochastic. The randomness could also be attributed to the underlying probabilistic mechanics of the atoms involved in the collective process. In either case, a stochastic approach to modeling the collective process can both resolve the factors that influence a given process and increase our understanding of the range of behaviors one can expect from a model.

We note the following ways we have observed the incorporation of stochasticity into coarse-grained models, each of which are described in more detail in the sections that follow:

1. Random selection of events from specific probability distributions
2. Deterministic selection of events from specific probability distributions
3. Emerging applications of other methods

### 2.2.1 Random Selection of Events from Specific Probability Distributions

Random selection of events from specific probability distributions is most frequently implemented in Monte Carlo models that utilize Maxwell-Boltzmann probability distributions. These include both standard Monte Carlo (Rollett and Manohar 2005) and kinetic Monte Carlo (Voter 2007) methods.

In these methods, one considers an ensemble of potential states into which the system can evolve. The selection of any given event is accomplished by developing a partition function based on the probability distribution function of interest. As noted, in Monte Carlo methods, the Maxwell-Boltzmann distribution is most commonly used, though other probability distributions can be used (Janssens et al. 2007). Once the partition function is created, the selection of events involves random numbers.

One of the best known applications of Monte Carlo methods is to the kinetic process of diffusion, where particle motions are modeled with a random, or stochastic, walk. Diffusion coefficients can then be obtained from a series of individual atomic jumps (Perondi et al. 1994).

In standard Monte Carlo, a system evolves through equilibrium (often metastable) states, with no knowledge of the transition states separating them. For this reason, time cannot be directly considered, although indirect conversion from Monte Carlo steps to physical time may be attempted by comparing to a known evolution law. While time can be difficult to access by this approach, one gains insight into the most favorable energetic states of the system. Among the many ways to execute this method is the Metropolis algorithm (Rollett and Manohar 2005; Janssens et al. 2007). In this algorithm, a site is selected at random, and then the energy change, $\Delta E$, is calculated for a state change from the starting state to a possible ending state. A random number between 0 and 1 is then compared to the transition probability $\exp(-\Delta E/k_B T)$ based on the Maxwell-Boltzmann distribution. In this probability, $k_B$ is the Boltzmann constant and $T$ is the temperature in Kelvin. If the random number is less than the transition probability, the event is accepted; otherwise it is rejected. This process is repeated over and over to evolve the system. There are different variants of the algorithm for faster execution and for conserved and non-conserved quantities. This type of algorithm has been used to study microstructure evolution extensively.

In kinetic Monte Carlo, a system evolves by consideration of the transition states rather than the equilibrium states. For this reason, time is accessible and one gains insight into the most likely evolution paths. This method typically follows a single implementation and is known as the kinetic Monte Carlo algorithm (Voter 2007). In this algorithm, a system is investigated for all the possible escape pathways that lead out of the current state. The energy change, $\Delta E$, of the transition state relative to the starting state is calculated for all the possible escape pathways. The transition rate $v_o \exp(-\Delta E/k_B T)$ is then calculated for each escape pathway. Note that in contrast to the transition probabilities of standard Monte Carlo, the kinetic Monte Carlo algorithm calculates a rate of successful transitions out of the current state by including a prefactor $v_o$ multiplying the transition probability. The prefactor $v_o$ is an attempt frequency that is typically of the order of the Debye frequency.

The cumulative transition rates, proportional to the partition function, are used to calculate the cumulative probability distribution, from which an event is selected using a random number. The average time for any one of the events to occur is inversely proportional to the cumulative transition rates, but since the probability of executing one of those events in that time interval must be selected from an exponential distribution, the elapsed time is usually selected as $t = -(1/k_{tot}) \ln(r)$, where $k_{tot}$ is the cumulative rate and $r$ is a random number between 0 and 1. This process is repeated over and over to allow the system to evolve by passing through transition states. This type of algorithm is extremely useful for systems with infrequent events.

In both of these cases, random numbers are an integral part of evolving the systems and are used to select events whose probabilities are based on the energetics of the processes of interest.

### 2.2.2   Deterministic Selection of Events from Specific Probability Distributions

Deterministic selection of events from specific probability distributions does not have a normal framework under which it is executed like the Monte Carlo methods do. Instead, these approaches typically use frameworks that are unique to the model and phenomena.

Since each implementation is unique, we briefly describe a basic approach, where local material property values are assigned in a stochastic manner but deformation proceeds based on a deterministic criterion. One such approach was used by Li and Anderson (2016) and is discussed briefly in Sect. 3.2.1. As noted in the introduction, strength at the nanoscale can be highly variable and can therefore be better represented by a strength distribution than a single value of strength (Uchic et al. 2004; Packard et al. 2010; Askari et al. 2015). These strength distributions can adopt some form of a probability distribution as a function of stress. One can use deterministic criteria such as those frequently used in crystal plasticity simulations (Roters et al. 2010) to determine if a given region has reached the condition for yield. But, instead of having each coarse-grained region of the model possess an identical strength value, one can randomly assign strength values as selected from the probability distribution of interest. Thus, strength is stochastically assigned according to a specific probability distribution. However, it is worth noting that we consider this approach to be deterministic selection in spite of the fact that values are randomly assigned. This is because the evolution law is governed by a deterministic comparison of values, and in most cases these values have been randomly assigned at the beginning and do not change during the course of the simulation.

In this way, these models sample physically relevant probability distributions of the property of interest. The models also select each of the events using standard deterministic approaches, and the coarse-grained models evolve accordingly.

### 2.2.3   Emerging Applications of Other Methods

The Monte Carlo methods are certainly the most common implementations of stochasticity in meso-scale models. However, the deterministic selection methods

noted above are increasing in frequency, and there are other methods that are more recently being used to describe the variability and randomness of different processes.

For example, stochastic finite element analysis has been around for some time, and one can use formulations such as polynomial chaos expansion to model the effects of uncertainty and/or variability in model parameters (Ghanem and Spanos 1990). Multi-fractal analysis (Lebyodkin and Estrin 2005) is another technique that uses probabilistic analysis to examine and describe collective phenomena. Random field theory enables modeling of random nonhomogeneous material properties at the meso-scale (Soize 2006; Guilleminot et al. 2009; Kouchmeshky and Zabaras 2010).

These are just a few examples of techniques that can describe stochastic aspects of collective phenomena at the meso-scale. These techniques differ from the two categories noted above for random and deterministic selection of events from specific probability distributions. In both of the categories above, one can point to specific, or discrete, processes that have occurred at any given step. In contrast, the methods noted here are more in line with deterministic methods that incorporate elements of stochasticity and do not deal with specific, or discrete, mechanisms. While different in nature, these methods have notable attributes that make them ideal for modeling some material behaviors and are likely to become more common.

## 3 Implementations of Stochastic, Coarse-Grained Models

This section discusses different implementations of stochastic, coarse-grained models for materials mechanics. We first provide a brief overview of the chapters included in this section of the handbook and the manner in which each implements the elements of stochasticity and coarse-graining. We follow up with a brief overview of other notable examples available in the literature.

### 3.1 Contributions of This Section

#### 3.1.1 Shear Transformation Zone Dynamics Modeling of Deformation in Metallic Glasses

Metallic glasses, also known as amorphous metals, are metallic alloys that have no long-range crystalline structure. As a result, their deformation differs significantly from that of crystalline materials. Metallic glasses exhibit glass-like flow at high temperatures and localized deformation, in the form of shear bands, at low temperatures. The time- and length-scales associated with these behaviors make these materials a challenge to model using atomistic methods or continuum mechanics. In ▶ Chap. 56, "Shear Transformation Zone Dynamics Modeling of Deformation in Metallic Glasses" Li and Homer examine the deformation behavior of metallic glasses using a coarse-grained, kinetic Monte Carlo algorithm. They coarse-grain the system on regions of material, known as shear transformation zones, in which atoms are shown to deform collectively. The kinetic Monte Carlo algorithm is used to evolve the system, where the energetics for the Boltzmann probabilities are based on models for the mechanics associated with shear transformation zones. In their

work they examine the process by which shear bands form, structural factors that influence shear banding, and the manner in which individual shear bands contribute to the overall deformation behaviors that are observed in metallic glasses.

### 3.1.2 Object Kinetic Monte Carlo: A Coarse-Grained Approach to Radiation Damage

Nuclear materials undergo complex processes as a result of the irradiation and accompanying thermal and mechanical conditions. They experience everything from radiation damage, void formation, fission gas generation, and swelling to phase changes, diffusion of species, and chemical reactions with cladding materials. Unfortunately, accurate modeling requires many of these processes to be considered simultaneously. This results in serious challenges when selecting an appropriate modeling technique. In ▶ Chap. 58, "Object Kinetic Monte Carlo (OKMC): A Coarse-Grained Approach to Radiation Damage" by Domain and Becquart examine the radiation damage in nuclear materials using an object kinetic Monte Carlo approach. They model different microstructural features, such as vacancies, voids, and interstitial clusters, as objects. The kinetic Monte Carlo algorithm is then used to control the evolution based on possible material processes. These processes include, among others, vacancy migration and recombination, which are selected from Boltzmann probabilities based on the respective energetics of the process. Using this model they examine the effect of impurities on the evolution of vacancies and ways to model these processes at different scales in combination with atomistic and mean-field approaches.

### 3.1.3 Kinetic Monte Carlo Modeling of Martensitic Phase Transformation Dynamics

Shape-memory alloys, a representative class of superelastic (pseudoelastic) materials, and other materials exhibiting martensitic phase transformations display an interesting ability to accommodate, or impose, deformation through displacive phase transformations. In these materials a high-temperature, austenitic, phase exhibits a different crystal structure than a low-temperature, martensitic, phase. Thus, when a region of the material changes phase, the surrounding material is forced to accommodate the difference in structure, which can be significant. In ▶ Chap. 57, "Kinetic Monte Carlo Modeling of Martensitic Phase Transformation Dynamics" by Chen examines the behavior of a shape-memory alloy using a coarse-grained kinetic Monte Carlo approach. Regions of material are coarse-grained to represent discrete regions that transform collectively. The kinetic Monte Carlo algorithm is then used to evolve the system using Boltzmann probabilities based on the energetics of the phase transformation and accompanying mechanics. The model is used to examine mechanically- and thermally-driven phase transformations, along with their accompanying hysteretic behaviors.

### 3.1.4 The Stochastic Nature of Deformation Twinning: Application to HCP Materials

Hexagonal close-packed (HCP) materials often accommodate deformation using a combination of dislocation plasticity and deformation twinning. Dislocation

plasticity follows well-known laws of initiation and propagation, whereas deformation twinning is less well understood. For example, twin formation appears to be random and is heterogeneous throughout the microstructure. Generating a model that can accurately capture the conditions for nucleation, twin variant selection, and subsequent growth of the twin is challenging. In ▶ Chap. 59, "The Stochastic Nature of Deformation Twinning: Application to HCP Materials" Beyerlein and Kumar examine twinning in HCP materials using a mean-field model that tracks coarse-grained features of a microstructure. Nucleation and growth of twins follow a deterministic law which instigates the various processes as determined from specific probability distributions. The model is used to show that stochastic selection of twin variants more accurately captures experimental observations than alternative approaches. The elements of stochasticity are important in accounting for the large variation of features that impact twinning in HCP metals.

## 3.2 Contributions in the Literature

There are also many notable contributions in the literature and we wish to highlight a few of those here.

### 3.2.1 Quantized Crystal Plasticity Modeling of Nanocrystalline Metals

One recent contribution focuses on dislocation plasticity in metallic alloys. Dislocation plasticity involves the motion and interaction of innumerable linear defects. The mechanics of motion of individual dislocations are well known, but the interaction of large numbers of dislocations remains a challenge. While many methods exist, Li and Anderson (2016) have developed a method where dislocation plasticity is accumulated in quantized amounts. Furthermore, the evolution of the dislocation plasticity is controlled by a deterministic evolution law, but the strengths of coarse-grained regions in the model are assigned from specific probability distribution functions. The model is used to capture several properties of dislocation plasticity that are unique to nanocrystalline metals.

### 3.2.2 Monte Carlo Finite Element Model for Strain Energy Controlled Microstructural Evolution

The formation and evolution of $\gamma'$ precipitates in nickel-based superalloys is a precipitation process that is strongly influenced by strain energy. To model the coalescence of layers, or rafts, of the $\gamma'$ precipitates, Gayda and Srolovitz (1989) developed a combined Monte Carlo-finite element model to include the strain energy in the Monte Carlo energetic calculations. Their model can account for externally applied stresses, surface tension, misfit, elastic inhomogeneity, elastic anisotropies, and temperature. The model shows good agreement with experimentally observed microstructures.

### 3.2.3   A Polynomial Chaos Expansion Based Molecular Dynamics Study for Probabilistic Strength Analysis of Nano-twinned Copper

To demonstrate how polynomial chaos can be used to represent the collective processes of interest in this section, we include this example from Mahata et al. (2016). In their work, they construct a polynomial chaos expansion to represent the influence of twin spacing and twin boundary width on strength in nano-twinned structures. Results from a small number of molecular dynamics simulations are used to construct the expansion, from which a larger variation of parameters is considered. The polynomial chaos expansion model is able to construct robust inferences of parameters that are only partially considered in the molecular dynamics simulations. This approach demonstrates emerging opportunities to incorporate the element of stochasticity into coarse-grained models.

### 3.2.4   Other Noteworthy Publications

Other noteworthy stochastic, coarse-grained models or models with similar approaches have been published. The following list is to illustrate a diversity of applications and methods and is not meant to be comprehensive. These include works by Guilleminot et al. (2009), Kouchmeshky and Zabaras (2010), Talamali et al. (2012), Soize (2012), Zhao et al. (2013), Askari et al. (2015), and Kondori et al. (2018).

## 4   Outlook and Open Challenges

Scientists may one day fulfill Einstein's wish of a unified field theory, but until that day, we must utilize the tools that capture the stochastic nature of materials.

Incorporating the element of stochasticity in coarse-grained models enables collective processes to be modeled in a way that captures their inherent variability, uncertainty, and probabilistic nature. These models are an important part of multi-scale modeling and bridge important length- and time-scales. It is expected that these approaches will fulfill an increasing need as we improve our understanding of quantum and atomistic processes but struggle to apply these to complex systems at the continuum scale. This means that there are a number of open challenges and opportunities that exist. A limited list is included here:

- There are numerous material processes that exhibit stochasticity and may benefit from a modeling approach that accurately captures their inherent nature. A brief list of these include fracture, fatigue crack growth, dislocation interactions with defects, martensitic transformations in steels, and recrystallization. Each of these processes is controlled by well-known energetic processes, but predicting exactly what will happen remains a challenge. Stochastic modeling approaches enable statistics and probabilities to be measured. As a result, new insight can be gained or energetic models can be improved. Furthermore, materials often undergo simultaneous mechanical deformation and kinetic coarsening of grains

and precipitates, both of which are somewhat stochastic processes. Coupled modeling of mechanical and kinetic evolution pathways will be required in order to accurately predict material behavior in many extreme environments.

- Different material processes exhibit different collective processes and probabilities that are best modeled by different approaches. One future area may be to develop coarse-grained, physics-based formulations and determine which modeling approaches are the most efficient or provide the most insight into a given process.

- While stochastic, coarse-grained models are not new, their use has been limited, and they appear to be gaining more attention as computational resources increase in power and capability. Newer applications, like polynomial chaos expansion, are being developed, and newer methods will certainly be discovered. The open challenge then is to determine new methods that allow processes to be modeled more efficiently or discover new methods that enable new material processes to be modeled in a stochastic manner that were not previously possible.

- One other direction that offers tremendous opportunity lies in the integration of these types of stochastic, coarse-grained models in a materials design framework. While the examples are focused on understanding and predicting material response under external stimuli, the models and tools can be further integrated or modified for materials design purposes. While uncertainty in design is gaining increasing interest, integrating variability in design and structure with predictive stochastic models will offer unprecedented opportunity in the design and manufacturing of high-performance materials.

The ultimate benefit of these modeling approaches will be proven by their use and ability to provide new insight into the materials that surround us. We hope the chapters that follow will aid and inspire new approaches and research that keep the community moving forward.

# References

Askari H, Maughan MR, Abdolrahim N, Sagapuram D, Bahr DF, Zbib HM (2015) A stochastic crystal plasticity framework for deformation of micro-scale polycrystalline materials. Int J Plast 68:21–33. https://doi.org/10.1016/j.ijplas.2014.11.001

Beden SM, Abdullah S, Ariffin AK (2009) Review of fatigue crack propagation models for metallic components. Eur J Sci Res 28(3):364–397

Bhattacharya K (2012) Microstructure of martensite: why it forms and how it gives rise to the shape-memory effect. Oxford University Press, Oxford

Born M, Einstein A (1971) The Born-Einstein letters. Macmillan, London

Christian JW (2002) The theory of transformations in metals and alloys. Pergamon, Amsterdam

Ellyin F (1997) Fatigue damage, crack growth and life prediction. Chapman & Hall, London. https://doi.org/10.1007/978-94-009-1509-1

Gayda J, Srolovitz DJ (1989) A Monte Carlo-finite element model for strain energy controlled microstructural evolution: "rafting" in superalloys. Acta Metallurgica Et Materialia 37(2):641–650. https://doi.org/10.1016/0001-6160(89)90248-4

Ghanem R, Spanos PD (1990) Polynomial chaos in stochastic finite elements. J Appl Mech 57(1):197–202. https://doi.org/10.1115/1.2888303

Guilleminot J, Soize C, Kondo D (2009) Mesoscale probabilistic models for the elasticity tensor of fiber reinforced composites: experimental identification and numerical aspects. Mech Mater 41(12):1309–1322. https://doi.org/10.1016/j.mechmat.2009.08.004

Hemminger J, Crabtree G, Sarrao J (eds) (2012) From quanta to the continuum: opportunities for mesoscale science. A report for the Basic Energy Sciences Advisory Committee, Department of Energy. https://www.osti.gov/servlets/purl/1183982/

Janssens KG, Raabe D, Kozeschnik E, Miodownik MA, Nestler B (2007) Computational materials engineering: an introduction to microstructure evolution. Academic Press, Boston

Kondori B, Amine Benzerga A, Needleman A (2018) Discrete shear-transformation-zone plasticity modeling of notched bars. J Mech Phys Solids 111:18–42. https://doi.org/10.1016/j.jmps.2017.10.010

Kouchmeshky B, Zabaras N (2010) Microstructure model reduction and uncertainty quantification in multiscale deformation processes. Comput Mater Sci 48(2):213–227. https://doi.org/10.1016/j.commatsci.2010.01.001

Kozin F, Bogdanoff JL (1989) Recent thoughts on probabilistic fatigue crack growth. Appl Mech Rev 42(11S):S121–S127. https://doi.org/10.1115/1.3152380

Lebyodkin MA, Estrin Y (2005) Multifractal analysis of the Portevin–Le Chatelier effect: general approach and application to AlMg and AlMg/Al$_2$O$_3$ alloys. Acta Mater 53(12):3403–3413. https://doi.org/10.1016/j.actamat.2005.03.042

Li L, Anderson PM (2016) Quantized crystal plasticity modeling of nanocrystalline metals. In: Weinberger CR, Tucker GJ (eds) Multiscale materials modeling for nanomechanics. Springer International Publishing, Cham, pp 413–440. https://doi.org/10.1007/978-3-319-33480-6_13

Mahata A, Mukhopadhyay T, Adhikari S (2016) A polynomial chaos expansion based molecular dynamics study for probabilistic strength analysis of nano-twinned copper. Mater Res Express 3(3):036501. https://doi.org/10.1088/2053-1591/3/3/036501

Miyamoto G, Iwata N, Takayama N, Furuhara T (2013) Variant selection of lath martensite and bainite transformation in low carbon steel by ausforming. J Alloys Compd 577:S528–S532. https://doi.org/10.1016/j.jallcom.2011.12.111

Packard CE, Franke O, Homer ER, Schuh CA (2010) Nanoscale strength distribution in amorphous versus crystalline metals. J Mater Res 25(12):2251–2263. https://doi.org/10.1557/JMR.2010.0299

Perondi LF, Elliott RJ, Barrio RA, Kaski K (1994) Tracer diffusion in lattices with double occupancy of sites. Phys Rev B 50(14):9868–9874. https://doi.org/10.1103/PhysRevB.50.9868

Provis JL (2015) Grand challenges in structural materials. Front Mater 2:31. https://doi.org/10.3389/fmats.2015.00031

Rollett AD, Manohar P (2005) The Monte Carlo method. In: Raabe D, Roters F, Barlat F, Chen LQ (eds) Continuum scale simulation of engineering materials. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 77–114. https://doi.org/10.1002/3527603786.ch4

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58(4):1152–1211. https://doi.org/10.1016/j.actamat.2009.10.058

Soize C (2006) Non-Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators. Comput Methods Appl Mech Eng 195(1–3):26–64. https://doi.org/10.1016/j.cma.2004.12.014

Soize C (2012) Stochastic models of uncertainties in computational mechanics. ACSE, Reston

Tadmor EB, Miller RE (2011) Modeling materials. Continuum, atomistic and multiscale techniques. Cambridge University Press, Cambridge

Talamali M, Petäjä V, Vandembroucq D, Roux S (2012) Strain localization and anisotropic correlations in a mesoscopic model of amorphous plasticity. Comptes Rendus Mecanique 340(4–5):275–288. https://doi.org/10.1016/j.crme.2012.02.010

Uchic MD, Dimiduk DM, Florando JN, Nix WD (2004) Sample dimensions influence strength and crystal plasticity. Science 305(5686):986–989. https://doi.org/10.1126/science.1098993

Voter AF (2007) Introduction to the Kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, Uberuaga BP (eds) Radiation effects in solids. Springer, Dordrecht, pp 1–23. https://doi.org/10.1007/978-1-4020-5295-8_1

Yip S, Short MP (2013) Multiscale materials modelling at the mesoscale. Nat Mater 12(9):774–777. https://doi.org/10.1038/nmat3746

Zhao P, Li J, Wang Y (2013) Heterogeneously randomized STZ model of metallic glasses: softening and extreme value statistics during deformation. Int J Plast 40:1–22. https://doi.org/10.1016/j.ijplas.2012.06.007

# Shear Transformation Zone Dynamics Modeling of Deformation in Metallic Glasses

# 56

Lin Li and Eric R. Homer

## Contents

## Abstract

A mesoscale shear transformation zone (STZ) dynamics model is presented to investigate the deformation behaviors of metallic glasses that span significant time and length scales. The modeling framework involves coarse-graining STZs, the fundamental deformation unit in metallic glasses, onto a finite element mesh and controlling the stochastic activation of these STZs using the kinetic Monte Carlo algorithm based on the energetics of the glass system. The combination of these two features allows simulating diverse deformation modes of metallic

L. Li (✉)
Department of Metallurgical and Materials Engineering, The University of Alabama, Tuscaloosa, AL, USA
e-mail: lin.li@eng.ua.edu

E. R. Homer
Department of Mechanical Engineering, Brigham Young University, Provo, UT, USA
e-mail: eric.homer@byu.edu

glasses at large time and length scales while providing a microscopic view of the process that dominates the behaviors. The adaption of the STZ dynamics framework to treat complex phenomena is discussed, including a detailed examination of the shear banding process, simulated contact mechanics, and an examination of the interplay of deformation and structural evolution via the incorporation of a free volume state variable. The chapter concludes with the challenges and future development of the STZ dynamics model.

# 1    Introduction

Metallic glasses exhibit deformation behaviors that are both interesting and challenging to fully characterize. As a result, a complete understanding of this new class of materials requires modeling techniques that can capture the relevant phenomena at the appropriate scales (Rodney et al. 2011; Schuh et al. 2007). The amorphous structures of metallic glasses appear deceptively homogeneous and isotropic when investigated with conventional experimental characterization techniques; the materials don't exhibit typical defect structures, such as dislocation, present in crystalline alloys. Despite the lack of experimental means to directly characterize flow defects, the state of the art in material modeling offers highly accurate methods, from density functional theory and molecular dynamics at atomistic scales all the way up to continuum theories for the deformation of metallic glasses and its connection with amorphous structures.

The deformation of metallic glasses, in particular, the low-temperature shear banding behavior is a typical multiscale phenomenon (Greer et al. 2013), occurring over several length and time scales illustrated in Fig. 1. At the atomic scale, the fundamental units of deformation are atomic rearrangements called shear transformation zones (STZs) (Argon 1979). The STZs are localized both in space, involving only a few tens of atoms, and time, spanning a few picoseconds, which have been captured and extensively studied by atomistic simulations (Falk and Maloney 2010; Rodney et al. 2011). The structural origin of STZs remains elusive, and yet it is believed that the inhomogeneous atomic packing configurations in the amorphous structure lead to a heterogeneous local inelastic response, linking to STZs (Cheng and Ma 2011; Ma and Ding 2016). Figure 1e displays a reconstructed atomic configuration of $Zr_{50}Cu_{45}Al_5$ metallic glasses by hybrid reverse Monte Carlo simulation, revealing a variety of packing clusters including the ideal icosahedron, distorted icosahedron, and face-centered cubic structure. Using molecular dynamics simulations, Ding et al. demonstrate that the regions densely populated with unstable clusters are elastically soft and more susceptible to be STZs (Fig. 1f). Such nanoscale inelastic heterogeneities have been recorded experimentally using dynamic atomic force microscopy (Fig.1b) and nanoindentation (Fig.1c). At the mesoscopic scale, the collective behaviors of STZs lead to various unique deformation phenomena in amorphous materials, such as spontaneous strain localization/shear banding, intermittent dynamics, and power-law distributed avalanches, which receive considerable attention from mesoscale modeling and experiments (Dahmen et al. 2009; Rodney et al. 2011). The connection of these deformation
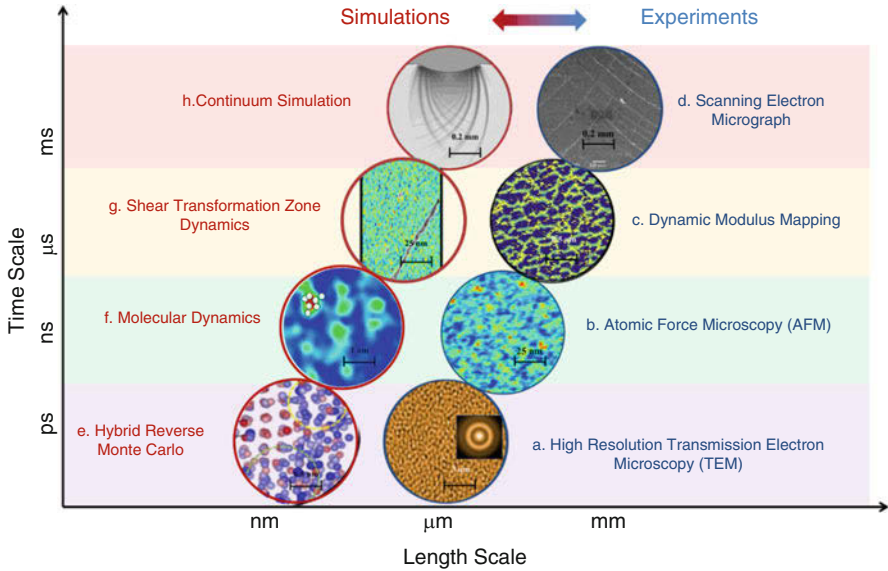
**Fig. 1** Multiscale features of the deformation behaviors and structural heterogeneities in metallic glasses. (**a**) The disordered atomic structure imaged by high-resolution transmission electron microscopy (HRTEM) (Ma and Zhang 2010). (**b**) The inelastic phase shift image of $Zr_{55}Cu_{30}Ni_5Al_{10}$ metallic glass with a correlation length $\sim$5 nm using atomic force microscopy (AFM) (Liu et al. 2011). (**c**) The elastic microstructure of $Zr_{58.5}Cu_{15.6}Ni_{12.8}Al_{10.3}Nb_{2.8}$ metallic glass with a characteristic length $\sim$100 nm characterized by dynamic modulus mapping on nanoindentation platform (Tsai et al. 2017). (**d**) Intersecting shear bands under the indenter of a Zr-based MG using scanning electron microscopy (SEM) (Su and Anand 2006). (**e**) The reconstructed atomic clusters of $Zr_{50}Cu_{45}Al_5$ MG by hybrid reverse Monte Carlo simulations (Hwang et al. 2012). (**f**) The atoms experienced shear transformation overlap on the participation ratio of soft modes in a $Cu_{64}Zr_{36}$ MG obtained by molecular dynamics (MD) simulations (Ding et al. 2014). (**g**) The simulated shear band formation of a $Cu_{64}Zr_{36}$ MG in the presence of elastic heterogeneity using a mesoscale shear transformation zone (STZ) dynamics simulations (Wang et al. 2018). (**h**) Simulated shear band pattern under indentation using a continuum model (Su and Anand 2006)

behaviors to the amorphous structure is beyond short-range orders; the atomic heterogeneities coordinate over a larger scale, translating from the nanometer-scale STZs to their organization into shear bands (Fig. 1g), which usually appear within a few milliseconds and reach a length on the order of a tenth of a micrometer. Finally, at the macroscopic scale, depending on the loading condition, either a single shear band forms as in tension test, or several shear bands form and interact, in case of confined plasticity as in indentation tests, as shown in Fig. 1d, h. The continuum modeling of the deformation behavior of metallic glasses has mostly relied on a flow rule accounting for the evolution of an internal state variable, the free volume, proposed by Spaepen (1977), relating the plastic strain rate to the state of stress and the history of deformation of the glass.

On the modeling front, atomistic simulations are critical in resolving the physics and mechanics associated with individual STZ activations, the nature of STZ-STZ interactions (Falk and Maloney 2010) as well as the incipient stage of shear band

nucleation (Şopu et al. 2017). Unfortunately, the atomistic simulations limit in both time and length scale to simulate the shear banding behaviors at engineering scales. Continuum approaches, on the other hand, can model deformation at engineering scales and provide an ideal comparison to experiments (Su and Anand 2006). However, the presumed constitutive laws can only exhibit phenomena they have been designed to capture, and often the deformation physics that must be introduced is not completely understood. As a result, a mesoscale modeling technique is an important inclusion in modeling the deformation behavior of metallic glasses across the entire spectrum (Rodney et al. 2011), contributing toward a complete understanding of structure-mechanical property relationships in metallic glasses (Hufnagel et al. 2016).

Modeling the amorphous plasticity at the mesoscopic scale, one employs a coarse-grained description of the STZs and accounts for the dynamics of these elementary STZ processes; by averaging out atomistic effects, one can access larger scales in the same way as dislocation dynamics describes crystal plasticity. Rodney et al. have summarized the key ingredients of mesoscale models for amorphous materials, which include a local activation/yield criterion of STZs, an elastic coupling between STZ and amorphous matrix like an Eshelby inclusion, the evolution rule of activation/yield criterion, and a dynamical rule that associates a time scale to the STZ activations (Rodney et al. 2011). Based on the different choices of these rules, the mesoscale models for metallic glasses fall into three categories. First, a depinning model, developed by Vandembroucq et al. (Baret et al. 2002), employs statistical distributions of yield stresses as well as the transformation strains for STZs, driving the evolution of STZs via internal stress arising from the accumulation of Eshelby fields and extremal dynamics. Second, a fluidity model, developed by Picard et al. (2002), uses a constant yield-stress criterion for STZs and incorporates the glassy dynamics based on Maxwellian viscosity through a distribution of characteristic transition rates. Third, an STZ dynamics model (Bulatov and Argon 1994a; Homer and Schuh 2009; Zhao et al. 2013), which is the focus of this chapter, uses an energy-based activation criterion for the STZs and a kinetic Monte Carlo algorithm to evolve the system through Boltzmann statistics.

In this chapter, we discuss the development of the STZ dynamics modeling framework and its applications in various aspects of metallic glass deformation, detailing the techniques used to bridge the relevant time and length scales. In addition, we examine the deformation physics elucidated by this method as well as the mechanics associated with shear banding behaviors.

## 2    STZ Dynamics Modeling Framework

The mesoscale STZ dynamics model treats STZs as the elementary deformation events, and the stochastic activation of the STZs leads to the formation of shear bands at large time and length scales (Homer et al. 2010; Homer and Schuh 2009). This initial development of STZ dynamics framework is inspired by the work of Bulatov and Argon (1994a, b, c). As with Bulatov, the modeling framework

employs two separate elements, coarse-graining and kinetic Monte Carlo (kMC) algorithm, which individually bridge the length and the time scales associated with deformation in metallic glasses. The coarse-graining method centers on the STZ, which consists of a cluster of atoms that exhibit the transient shearing motion consistently in the same manner (Maloney and Lemaître 2004, 2006; Rodney and Schuh 2009; Srolovitz et al. 1983). Then a simulated metallic glass is represented by a system of potential STZs. The coarse-graining enables more efficient sampling of larger system sizes. Meanwhile, to simulate longer system times more efficiently, the transient STZ activation is considered as a transition state between the initial and final equilibrium configurations. Thus, transition state theory (TST) and the kMC algorithm can be employed as long as knowledge of the energetic landscape, including the transition states, is available.

## 2.1 STZ Coarse-Graining with Finite Element Mesh

The STZ coarse-graining is accomplished by replacing the cluster of atoms that represents a potential STZ with a feature of finite element mesh, as illustrated in Fig. 2 (Homer and Schuh 2010, 2009). In this process, three criteria are imposed for proper representation of an STZ using a finite element mesh. First, the coarse-grain representation should approximate the shape of an STZ, which is believed to be roughly spherical. Second, the finite element representation of the potential STZs should allow them to overlap, since the STZ is a transient event and the atoms involved would never be restricted to participate in only one potential STZs. In other words, for a given element, it can participate in multiple STZ activations. Third, the coarse-grained STZ should accurately capture the analytical solution of an Eshelby inclusion (Eshelby 1957). This is supported by the original STZ theory paper, in which Argon modeled the STZ as an Eshelby inclusion (Argon 1979).
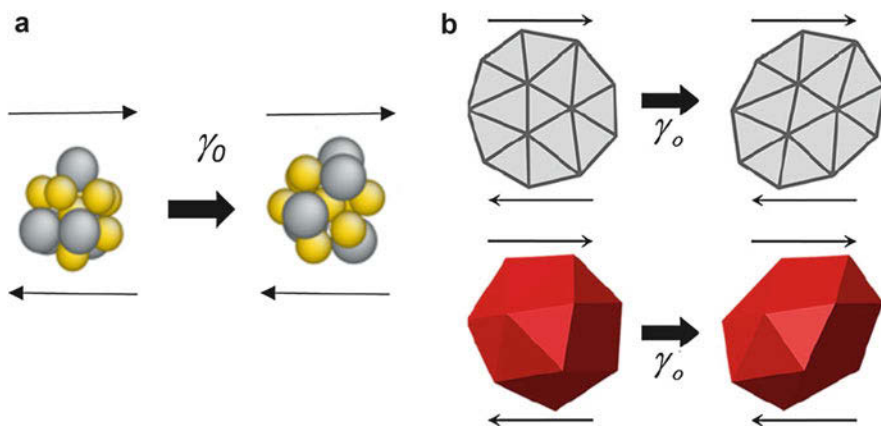


**Fig. 2** Coarse-graining of an STZ (**a**) using features of a finite element mesh in (**b**) 2D or 3D. (Figure adapted with permission from Homer and Schuh (2009, 2010))
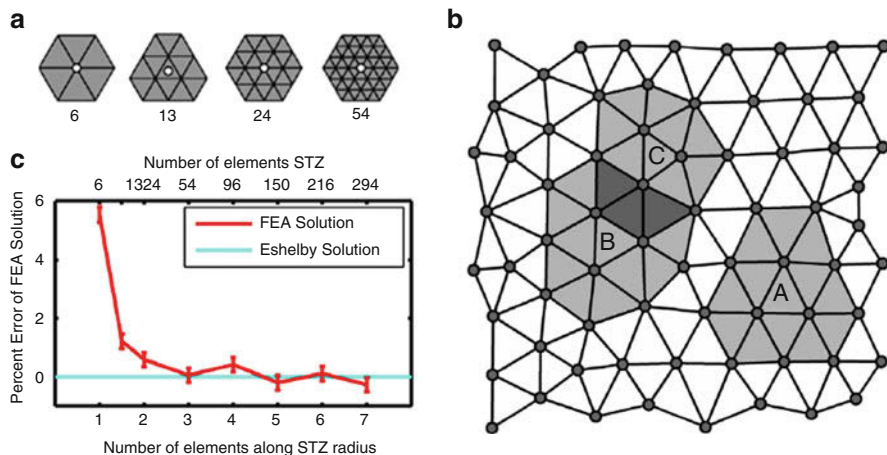
**Fig. 3** 2D coarse-graining process. (**a**) Representation of possible STZ definition on the triangular mesh. (**b**) Irregular triangular mesh with 13-element potential STZs highlighted to show how the individual elements in the mesh can be activated by different STZs. (**c**) The accuracy of STZ coarse-graining evaluated by the strain energy difference between finite element method and the Eshelby solution as a function of the size of the STZ (Figure adapted with permission from Homer and Schuh (2009))

Following the three criteria, the implementation of coarse-graining STZs onto 2D and 3D finite element meshes has been achieved (Homer and Schuh 2010, 2009). In 2D, a single STZ is represented by a node and all the surrounding elements or an element and all the surrounding elements on an irregular triangular mesh, shown in Fig. 3a. This representation ensures that the shapes of potential STZs are roughly equiaxed. Furthermore, in this representation, the STZs comprise more than one single element, so that elements in the mesh will be able to participate in multiple STZs. As illustrated in Fig. 3b, three potential STZs, each of 13 elements, are highlighted on an irregular triangular mesh. At any given time step, the elements in the overlap region (between potential STZs B and C) can participate in either event. Finally, the accuracy of the representation is evaluated by comparing to the Eshelby solution for shearing of a long circular fiber in a matrix (plane strain). The percent error of the calculation relative to the Eshelby solution (based on the total system strain energy) is plotted in Fig. 3c as a function of the size of the STZ relative to the mesh. As the results show, convergence is achieved quite rapidly, with STZs containing 13 or more elements exhibiting about 1.5% error or less (Homer and Schuh 2009).

In 3D, a collection of 20–30 tetrahedral elements that all share one common node provides a consistent approximately spherical STZ, as illustrated in Fig. 2b. In this definition, STZs may overlap. When compared with the Eshelby solution, the quadratic tetrahedral element-based STZs are found to provide higher accuracy with the elastic strain energy having 2% error (Homer and Schuh 2010).

The use of a finite element mesh not only enables a coarse-grained description of STZ as a transient flow defect in metallic glasses but also provides flexibility

for further development to incorporate emerging characteristics of STZs. First, one might define an ensemble of STZs with different characteristic volumes. In principle, this could be achieved based on the local size of the elements included in each potential STZ. In this chapter, a single value of volume is assigned to all the STZs in the mesh, but this value could vary with respect to strain rate (Harris et al. 2016) (see Sect. 3.2). Second, the size of STZ activation volume could be related to the level of glass relaxation and damage (Albaret et al. 2016; Boioli et al. 2017) and loading states (Fusco et al. 2010). Adaptive meshing could be used to capture the dynamic evolution of STZ volumes. Third, the elastic response of the system to the STZ transformation strain requires only the use of a linear elastic finite element solver. The use of finite element mesh and finite element solver enables just about any set of boundary conditions regularly used in finite element analysis to be incorporated into the STZ dynamics framework.

## 2.2  STZ Activation Rate

To accomplish the dynamics for longer time scale, the STZ dynamics model considers the coarse-grained STZ activation as a transition state between the initial and final equilibrium configurations, as illustrated in Fig. 4a. The stochastic activation of the STZs leads to deformation on longer timescales. The STZs are thermally activated, and the activation rate is proportional to the Boltzmann probability that the system overcomes the activation barrier between the initial and final equilibrium configurations, defined as
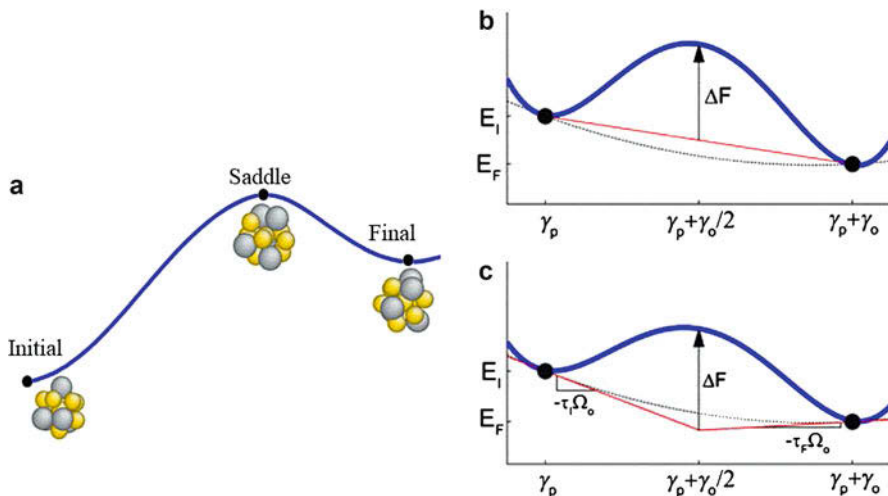


**Fig. 4** (**a**) Illustration of the potential energy landscape and the associated STZ configurations at the initial state, saddle point, and final state. (**b**) Illustration of the traditional approach of identifying the activated state of an STZ. (**c**) The energy landscape model proposed by Bulatov and Argon (1994a). (Figure (**b**) and (**c**) reproduced with permission from Homer et al. (2010))

$$\dot{s} = v_o \exp\left(-\frac{\Delta G}{k_B T}\right) \tag{1}$$

where the prefactor $v_o$ is of the order of the Debye frequency, $\Delta G$ is the activation energy barrier, T is the absolute temperature, and $k_B$ is the Boltzmann constant.

In order to calculate the activation energy barrier for a given transition, one must have knowledge of the transition itself. A traditional approach uses the energy change to model the activation energy, $\Delta G$, by adding a barrier of fixed height, $\Delta F$, to the average of $E_I$ and $E_F$, as illustrated in Fig. 4b. This approach satisfies detailed balance for the reaction because a forward transition traverses the same activated state as the reverse transition.

$$E^I + \Delta G_{I \to F} = E^F + \Delta G_{F \to I} \tag{2}$$

with $\Delta G_{I \to F} = (E^F - E^I)/2 + \Delta F$ and $\Delta G_{F \to I} = (E^I - E^F)/2 + \Delta F$. Unfortunately, this traditional approach of calculating $\Delta G$ requires calculation of the energy in the final state, which is computationally expensive for a large number of possible transitions in metallic glasses.

Bulatov and Argon have provided an alternate formulation for $\Delta G$ by exploiting the fact that the strain energy associated with shearing an STZ is a quadratic function of the transformation strain $\gamma_0$ (Bulatov and Argon 1994a). This quadratic variation in energy, shown as the dashed line in Fig. 4c, can predict the energy in the final state without explicit calculation. Bulatov and Argon simply extrapolate from the initial state, using the slope at that point, to the midpoint of the transition and then add the fixed barrier height, $\Delta F$. The energy difference between the traditional approach and that of Bulatov and Argon is very small (Bulatov and Argon 1994a; Homer et al. 2010). For more detailed explanation, including the requirement of detailed balances, please refer to Homer et al. (2010).

The STZ dynamics framework uses the Bulatov and Argon model and defines the activation energy barrier as

$$\Delta G = \Delta F - \frac{1}{2}\tau \cdot \gamma_o \cdot \Omega_o \tag{3}$$

where the intrinsic barrier height for the reaction, $\Delta F$, is biased by the local shear stress $\tau$, which is obtained by volume averaging the stress over the elements which comprise each potential STZ. The activation volume, $\gamma_0 \Omega_0$, of the STZ is comprised of the transformation strain increment associated with an STZ transformation, $\gamma_0$, and the volume of the STZ, $\Omega_0$. Argon developed a model for $\Delta F$ (Argon 1979), given as

$$\Delta F = \left[\frac{7 - 5v}{30(1 - v)} - \frac{2(1 + v)}{9(1 - v)}\beta^2 + \frac{1}{2\gamma_o} \cdot \frac{\hat{\tau}}{\mu(T)}\right] \cdot \mu(T) \cdot \gamma_o^2 \cdot \Omega_o \tag{4}$$

where the three terms in the brackets define the strain energy associated with shearing of the STZ, the strain energy associated with a temporary dilatation of

the STZ to allow the atoms to rearrange, and the frictional energy associated with the free shearing of the atoms over one another. In the equation of $\Delta F$, $\nu$ is Poisson's ratio, $\beta$ is a ratio of shear to dilatation (usually taken as 1), $\widehat{\tau}$ is the peak interatomic shear resistance between atoms, and $\mu(T)$ is the temperature-dependent shear modulus.

The use of Bulatov's and Argon's energy model allows the STZ dynamics model to explore a large number of transitions without calculating the energy of the final state, resulting in significant computational saving. However, a fixed activation energy barrier would be hard to represent the complex non-equilibrium states in metallic glasses that contain many-body interactions and strong disorder. Methods such as the nudged-elastic band (NEB) (Boioli et al. 2017; Xu et al. 2017) or the activation-relaxation technique (ART) (Fan et al. 2014; Malek and Mousseau 2000; Rodney and Schuh 2009) can be used in atomistic simulations to explore the potential energy landscape and find the exact activation energy barrier from any given equilibrium state. Widely distributed activation energies, depending largely on the processing history, are usually obtained (Rodney and Schuh 2009; Rodney et al. 2011). These atomistic energy barrier search methods are computationally intensive and do not readily translate into mesoscale models. The development of activation energy functional based on the atomistic energy barriers, or their associated features, would be beneficial to enrich the STZ dynamics model.

### 2.2.1 STZ Activation Rate in 2D

The expression for $\Delta G$ given in Eq. 3 defines the energy barrier for an STZ to shear in one direction. Since we are interested in calculating the range of barriers associated with shearing an STZ in any direction in space, the shear stress associated with each unique shear direction must be identified and enumerated.

In 2D, the shear stress for each unique shear direction around a circle can be evaluated using a Mohr's circle construct, which gives the shear stress along any direction of the circle as

$$\tau = \tau_{\max} \sin(\theta) \tag{5}$$

where $\theta$ is the angle to the stress state with stress $\tau$ and which is measured relative to the stress state with the highest principal stress. One can then integrate all shear directions by integrating $\theta$ over the interval $(0°, 360°)$. By combining Eqs. 1, 3, and 5, the integral STZ activation rate becomes

$$\dot{s} = \frac{v_o}{2\pi} \cdot \exp\left(-\frac{\Delta F}{k_B T}\right) \cdot \int_0^{2\pi} \exp\left(\frac{\tau_{\max} \cdot \sin(\theta) \cdot \gamma_o \cdot \Omega_o}{2k_B T}\right) d\theta \tag{6}$$

which evaluates to a modified Bessel function of the first kind, of order zero

$$\dot{s} = \frac{v_o}{2\pi} \cdot \exp\left(-\frac{\Delta F}{k_B T}\right) \cdot I_o\left(\frac{\tau_{\max} \cdot \gamma_o \cdot \Omega_o}{2k_B T}\right) \tag{7}$$

This particular form of the STZ activation rate is convenient because the analytical solution gives the rate for shearing an STZ in any direction in two dimensions with only one function evaluation. One can then use this integral rate in the kMC implementation to determine which STZs are likely to be activated.

### 2.2.2 STZ Activation Rate in 3D

The evaluation of the STZ activation rate in 3D is more complex than the 2D case due to the larger set of possible shear planes and shear directions, as well as the need to only evaluate unique combinations of shear planes and directions. In a generalized form, the integral activation rate can be defined as

$$\dot{s} = v_o \cdot \exp\left(-\frac{\Delta F}{k_B T}\right) \cdot \iiint\limits_{g \,\in\, G} \exp\left(\frac{\tau\,(\sigma,\, g) \cdot \gamma_o \cdot \Omega_o}{2 k_B T}\right) dg \qquad (8)$$

where $g$ is the orientation of any shear plane-shear direction combination belonging to the set $G$ of all unique combinations of shear planes and shear directions. The integral is three dimensional because the specific orientation of a shear plane and shear direction requires three parameters. The shear stress of that orientation $g$ is defined as $\tau(\sigma,\, g)$ to denote the fact that the triaxial stress state that exists in a given STZ must be transformed by $g$ to obtain the shear stress for that given shear plane and shear direction. No analytical solution to the integral in Eq. 8 could be found. The integral is numerically evaluated and tabulated for rapid recall during the modeling process while maintaining an error less than 0.01%. Due to the complexity of this calculation, the details are not discussed here but are available in (Homer and Schuh 2010).

### 2.2.3 STZ Activation Rate with Excess Free Volume as a Local State Variable

The initial implementation of the STZ dynamics framework used a fixed $\Delta F$ for all STZ events. Potentially important effects related to the glass state, e.g., level of glass relaxation and damage, have not been considered. To solve this limitation, local state variables can be included as part of the energetics that describe STZ activations. The purpose of the state variables is to (1) incorporate local activation energy fluctuation that in broad agreement with the widely distributed activation energies of metallic glass and (2) capture the evolution of the structure beyond the redistribution of stress and strain when an STZ is activated. One could choose from a range of state variables, such as atomic stress and strains, topological or chemical order, free volume, and fictive temperature. In the current STZ dynamics framework, a state variable of "free volume" has been implemented (Li et al. 2013, 2014).

The implementation of free volume is based on Argon's original definition of the STZ, in which he includes free volume as a state variable to capture the structural evolution of the system. Particularly, in our adaptation of the STZ dynamics framework, excess free volume, $f_v$, is defined as a normalized quantity where $f_v = 0$ corresponds to no excess free volume above the average polyhedral volume in a

dense random hard sphere glass, while $f_v = 1$ is an upper bound corresponding to a state where an STZ can be activated without accumulating any extra free volume.

The excess free volume influences the energy barrier for STZ activation, and the fixed barrier is modified as

$$\Delta F_{STZ}(f_v) = \Delta F_{shear} + \Delta F_{v0} \cdot g_{stz}(f_v) \tag{9}$$

where $\Delta F_{shear}$ captures the strain energy associated with shear (not dependent upon excess free volume) and $\Delta F_{v0}$ captures the strain energy associated with dilatation and friction of the atoms sliding over each other (dependent upon excess free volume). Equation 9 essentially alters Eq. 4 by recognizing that the first term in the bracket of Eq. 4 is only dependent on shear but not on the magnitude of excess free volume, whereas the last two terms in the bracket of Eq. 4 are dependent upon the magnitude of excess free volume. Furthermore, $\Delta F_{STZ}$ is smaller when greater excess free volume exists since the STZ needs to dilate less and the friction will be lowered. This change in the energy is captured by the function $g_{stz}$, which lowers the activation energy barrier as the excess free volume is increased.

Following a given STZ activation, the excess free volume within the activated STZ is increased since it is believed that the atoms are not able to immediately return to the original magnitude of excess free volume (Li et al. 2013).

In parallel to the activation of STZs, a competing process is introduced as the diffusive rearrangement (and destruction) of excess free volume to capture the effect of structural relaxation in metallic glasses. Following Argon's original model, the rate of diffusive rearrangement is given as

$$\dot{s}_D = (1 - f_v)\, v_D \exp\left(-\frac{\Delta G_D(f_v)}{k_B T}\right) \tag{10}$$

where $\Delta G_D(f_v)$ is the activation energy barrier for diffusive rearrangement, which is dependent upon the current magnitude of excess free volume. Higher excess free volume has a lower energy barrier given that it is farther from the equilibrium state. The quantity $(1 - f_v)$ reflects a decrease of available atomic sites for free volume diffusion as $f_v$ increases. The prefactor $v_D$ for the diffusive rearrangements is of the order of the Debye frequency.

It is noteworthy that the implementation of excess free volume and most other state variables at mesoscale remains phenomenological (Rodney et al. 2011). Metallic glasses do not have structural defects found in crystalline materials, such as dislocations and grain boundaries. The definitions of structural defects in disordered materials are not unique and would require some phenomenological presumptions and fitting parameters. A fundamental understanding is still lacking on the dynamics of inherent glassy structure and its connection with the properties of glasses such as aging or rejuvenation (Fan et al. 2017). It remains difficult to develop parameter-free theories based on defects. Atomistic simulations and experimental measurements at microscopic scale would advance the development of "defect"-level theories and their implementation into the mesoscale model.

## 2.3    Kinetic Monte Carlo Implementation

### 2.3.1    kMC with STZ Activation

Upon calculation of the STZ activation rate, the kMC algorithm (Voter 2007) is used to evolve a system consisting of an ensemble of STZs. In the system, each STZ may experience different local temperature, stress state, and local state when state variables are incorporated. The kMC algorithm proceeds according to the following steps, which are repeated for every transition:

1. Calculate and form a list of activation rates, $\dot{s}_i$, for each of the $i = 1, \ldots, N$ STZs in the ensemble, based on the current state of the system.
2. Calculate the normalized rate $\eta_i$ for each STZ via dividing the individual activation rate by the cumulative activation rate, $\eta_i = \dot{s}_i / \dot{s}_T$, for all STZs. The sum over the normalized transition rates is equal to one, i.e., $\sum_i \eta_i = 1$.
3. Generate two random numbers, $\xi_1$ and $\xi_2$, uniformly distributed on the interval (0, 1).
4. Update the elapsed system time with the residence time for the current configuration calculated according to $\Delta t = - \ln \xi_1 / \dot{s}_T$.
5. Select a single STZ by first defining the cumulative fraction of STZ rates up to and including the rate of STZ $j$ by $H_j = \sum_{i=1}^{j} \eta_i$, and then use the random number, $\xi_2$, to find the STZ which satisfies $H_{k-1} < \xi_2 \leq H_k$. When listed in a successive faction, $\xi_2$ falls on the subinterval in the list of normalized STZ rates, as illustrated in Fig. 5.



**Fig. 5** Schematic of the kinetic Monte Carlo STZ selection procedure. (**a**) How the random number $\xi_2$ can be used to select a single STZ for activation from a list of normalized individual STZ rates, $\eta_1, \eta_2, \eta_3, \ldots, \eta_i$. (**b**) The determination of the overlap, $\eta'$, between $\xi_2$ and $\eta_j$, which selects the angle of shear of the STZ. (Figure reproduced with permission from Homer and Schuh (2009))
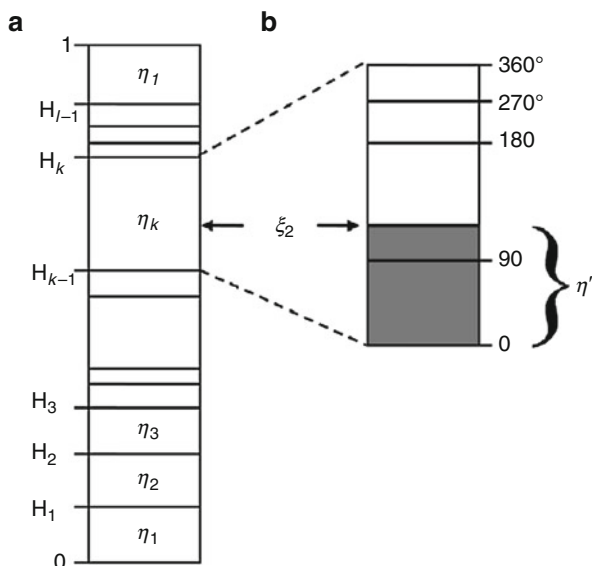
6. To select the angle at which to shear the STZ, we first define the value $\eta' = \xi_2 - H_{k-1}$, which represents the magnitude by which $\xi_2$ overlaps the subinterval of the selected STZ, as shown in Fig. 5. The overlap, $\eta'$, is then used to determine the integration limit and further the angle of shear in real space. The detailed explanation of angle selection can be found in Homer and Schuh (2009) for 2D and in Homer and Schuh (2010) for 3D.
7. Apply the appropriate transformation strain to the selected STZ.
8. Calculate the stress and strain distributions and state-variable evolution resulting from the new configuration.

The kMC algorithm can be repeated for an arbitrary number of STZ operations and is efficient because every iteration guarantees a transition. The stochastic nature of the processes will produce a realistic outcome as long as the rate law of the individual event is correct.

### 2.3.2   kMC Algorithm with Competing Processes

The kMC algorithm can easily be adapted to incorporate another type of processes competing with STZ activation to evolve the systems. For instance, after incorporating excess free volume as a state variable, a diffusive rearrangement process is introduced in competition with STZ activation. Upon implementation, at a given kMC step, the transition rates of diffusive rearrangement process will be included in addition to STZ activation to make a list of activation rates in Steps 1 and 2 described in Sect. 2.3.1. In Step 5, either diffusive rearrangement or STZ activation is selected, depending on the subinterval $\xi_2$ falls on in the list of the normalized transition rates. And thus, the two possible processes are exclusive; in each kMC increment, only one of them will be selected. The addition of competing processes is explained in more detail in (Li et al. 2013).

## 2.4   Summary of STZ Dynamics Framework

The STZ dynamics model simulates the stochastic activation of coarse-grained STZs, their elastic interaction leading to organization and accumulation of STZs forming shear bands at large time and length scales. The application of the modeling framework requires several steps to be followed. First, a 2D or 3D finite element mesh is defined to match the geometry of the model material being simulated. Second, potential STZs are mapped onto the finite element mesh based on the coarse-graining criteria discussed in Sect. 2.1. Third, a set of state variables can be assigned on the finite element mesh, influencing the material model and kMC algorithm, as discussed in Sect. 2.2.3. Fourth, implement the kMC algorithm and repeat the following steps:

1. Determine which STZ should be selected for activation, and which shearing angle should be applied, based on the current system state.

**Table 1** Material properties commonly employed by the STZ dynamics framework

| Property/variable | Value |
|---|---|
| Temperature-dependent shear modulus, $\mu(T)$ | $-0.004$ [GPa K$^{-1}$] $\times$ T $+$ 37 [GPa] |
| Poisson's ratio, $\nu$ | 0.352 |
| Debye temperature, $\theta_D$ | 327 K |
| Fixed activation energy barrier, $\Delta F(T)$ | $0.822 \times 10^{-29}$ [J Pa$^{-1}$] * $\mu(T)$ |
| STZ activation volume, $\Omega_o$ | 2.0 nm$^3$ |
| STZ strain, $\gamma_o$ | 0.1 |

2. Impose a characteristic transformation strain to the elements belonging to the selected STZ according to the selected shearing angle.
3. Use finite element analysis to determine the response of the system to the imposed transformation strains.
4. Update the current system state, including stress, strain, and any functional material properties, to reflect the response to STZ activation.

These last four steps involving the kMC algorithm are repeated many times in succession to determine the evolution of the system.

The key material properties and simulation variables used by many of the published STZ dynamics papers are listed in Table 1. The attempt frequency $\nu_o$ is taken as the Debye frequency, which can be calculated from the Debye temperature $\theta_D$. The variables $\mu(T)$, $\nu$, and $\theta_D$ have values for the commonly studied Vitreloy 1 with composition $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$ and are obtained from Johnson and Samwer (2005) and Wang et al. (2011b), respectively. Rather than using the complex form of the fixed barrier height in Eq. 4, we reduce $\Delta F$ to a simple functional form that is dependent upon the shear modulus. This functional form is also in line with the cooperative shear model proposed by Johnson and Samwer (Johnson and Samwer 2005). The STZ volume is in the range commonly reported in the literature (Zink et al. 2006), and the STZ strain is equal to the commonly accepted value (Schuh et al. 2007). This list of variables is intentionally kept short to simplify the model and obtain an intended response.

# 3 Applications of STZ Dynamics Model

The STZ dynamics modeling framework provides an opportunity to study many different aspects of metallic glass deformation. Since its original development, the STZ dynamics framework has been adapted for different implementations, including contact mechanics (Packard et al. 2010; Wang et al. 2015) and state-variable free volume evolution (Li et al. 2013; Wang et al. 2015), and for metallic glass matrix composites (Hardin and Homer 2015). In the following section, three applications are chosen to demonstrate:

1. The general behaviors of the STZ dynamics model and the corresponding spatial and temporal correlation of STZ activity that underlies various modes of deformation
2. The physics that control the low-temperature shear banding behavior and how the strain rate influences the shear banding process
3. How nanoindentation can be studied using the STZ dynamics modeling framework to gain insight into nanoscale strengthening in metallic glasses

## 3.1   General Behaviors and STZ Correlations

Metallic glasses exhibit a great variety of deformation behaviors, depending upon conditions (Schuh et al. 2007). The STZ dynamics model is able to capture the general MG behaviors, showing homogeneous deformation at high temperature and localization deformation into the shear bands at low temperature and high stress. A representative 2D model response is illustrated in Fig. 6a. The 3D model can capture the general behaviors as well (for details, please refer to (Homer and Schuh 2010)).

The model response over a range of conditions is well represented by the deformation map for simulation cells subjected to a range of applied stress at various temperatures. An example of deformation map for the 2D model is displayed in Fig. 6b. The regions of homogeneous deformation and inhomogeneous deformation are shaded. In addition, the steady-state strain rate is measured from each simulation, and contours of constant strain rate are overlaid on the map for rates ranging from $10^{-10}$ to $1$ $s^{-1}$. The shading inside each data point presents the strain rate sensitivity. At high temperature, as the stress is increased, the strain rate sensitivity decreases from unity toward zero, reflecting a transition from Newtonian flow to non-Newtonian flow. These are the rheological behaviors that metallic glasses exhibit at high temperature. At low temperature, when stress is low, the strain rates are lower than $10^{-10}$ $s^{-1}$, which we consider to be "elastic" deformation as the inelastic behaviors would be too slow to be captured in experiments. When the stress is high at low temperature, the strain rate sensitivity is extremely low; in other words, the flow stress is nearly the same for various strain rates. This is a consequence of the formation of shear bands, the details of which will be discussed in Sect. 3.2. The deformation map compares favorably with experimental deformation map, in that it captures the basic features of metallic glass deformation (Schuh et al. 2007).

Underlying the diverse deformation behaviors is the different spatial and temporal correlation of STZ activity. Analysis of the 2D simulations of STZ correlations provides significant insight, represented by the time-dependent radial distribution functions (TRDFs), given as

$$g\left(r, j\right) = \frac{n\left(r, j\right)}{q(r)} \tag{11}$$
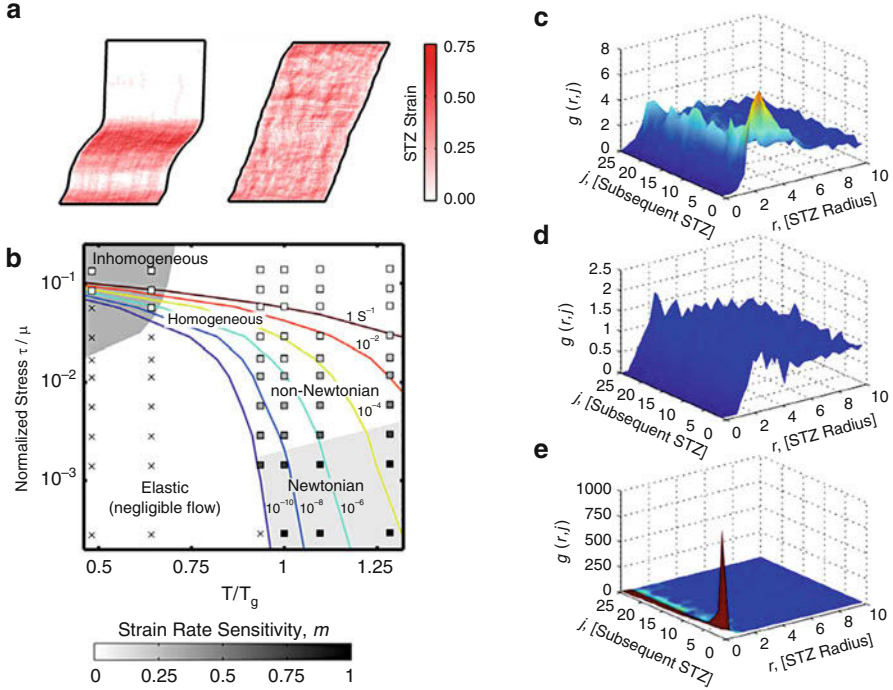
**Fig. 6** (**a**) The representative responses of the STZ dynamics model at high and low temperatures in 2D. (**b**) Deformation map for Vitreloy 1 constructed from data obtained by 2D STZ dynamics simulations. The STZ correlation behaviors represented by the TRDFs of STZ activation, where the three behaviors and their corresponding conditions are (**c**) nearest-neighbor STZ activation (high stress and low temperature), (**d**) independent STZ activation (high stress and high temperature), and (**e**) self-STZ activation (low stress and any temperature). (Figures adapted with permission from Homer et al. (2010); Homer and Schuh (2009))

where $n(r, j)$ is constructed by binning the number of sequential activations as a function of radius $r$ and time step $j$ and $q(r)$ is a normalization quantity in each bin with the size $dr$, defined as

$$q(r) = \begin{cases} 1 & if \ r \leq \frac{1}{2\pi r dr \rho_{STZ}} \\ 2\pi r dr \rho_{STZ} & if \ r > \frac{1}{2\pi r dr \rho_{STZ}} \end{cases} \tag{12}$$

where $\rho_{STZ}$ represents the overall density of STZ activations, i.e., the total number of STZ activations per unit area. The TRDFs' functions indicate the likelihood of shearing an STZ at nearby position and after a certain number of steps relative to a given STZ activation; magnitudes less than 1 are less likely to occur at a given position and time than if it occurs randomly throughout the simulation cell, and magnitudes greater than 1 are more likely to occur at a given position and time than if it occurred randomly throughout the simulation cell.

The TRDF reveals three basic types of behavior that manifest under different combinations of applied stress and temperature as shown in Fig. 6c–e.

- Nearest-neighbor STZ activation, which is observed for simulations at low temperature and high stress in Fig. 6c. This behavior is characterized by an early, broad peak, spanning $r = 1$–$5$, centered between 2 and 3, indicating the preference for nearest-neighbor activation. This correlated behavior is the source for the shear localization that underlies the macroscopic shear bands observed in experiments.
- Independent STZ activation, which occurs under conditions of high applied stress and high temperatures in Fig. 6d. In this behavior, the TRDF shows no preference for reactivation of STZs atop the first one, since $g(r, j) \sim 0$ at r < 1. Furthermore, the tendency for activation of neighboring STZs is lost; there is no longer a discernible peak in the TRDF. There is no noticeable correlation between STZ activations. As expected, the additional thermal energy cancels the effect of stress concentration that might otherwise cause shear localization. Consequently, the uncorrelated STZ activation leads to homogeneous deformation.
- Self-STZ activation, which dominates at low applied stress and any temperature. As illustrated in Fig. 6e, the TRDF exhibits an extremely pronounced and sharp peak at $r = 0$ and for early time ($j < 4$). The spatial extent of the peak is limited to $r \leq 1$, indicating a large preference for a second STZ activation atop the first. The self-STZ activation is linked to the elastic regime. At low temperatures and low stresses, there is an insufficient tendency for a single STZ to trigger nearest-neighbor activations; thus, the most likely response of the system is for each STZ activation to be instantaneously reversed.

The STZ dynamics model can not only capture the MG deformation behaviors at the macroscopic level, matching the experimental behaviors, but also provide insights at a microscopic level on how STZs interact with one another and how their collective operation leads to the deformation on a macroscopic level. These types of studies demonstrate the strength of a mesoscale model that successfully coarse-grains a process and determine the transitions that control the evolution of the system.

## 3.2 Shear Banding Process at Low Temperature

Among the diverse deformation modes exhibited by metallic glasses, the low-temperature shear banding behavior is of the greatest interest. The limited ductility due to the formation of a catastrophic shear band before failure is the primary issue that hinders the wide application of metallic glass as a structural material (Greer et al. 2013; Schuh et al. 2007). The STZ dynamics model can provide modeling details into the formation of the shear band, which remain unresolved due to the difficulty in accessing the appropriate time and length scales experimentally.

One can study the shear localization process into one shear band in detail (Homer 2014). Snapshots of a 3D simulation cell subject to constant strain rate, uniaxial tension test are shown in Fig. 7. Examination of the simulation results reveals five different stages in the microscopic processes:

(I) Purely elastic, with no STZ activity
(II) STZ clustering, where correlated STZ activations lead to the formation of clusters
(III) Growth following nucleation of a shear band, where all STZ activity transitions from being distributed throughout the simulation cell to being concentrated in the shear band
(IV) Relaxation thickening, which is manifest by the continued thickening of the shear band while the stress is still dropping even after it has propagated across the simulation cell
(V) Flow thickening, which is indicated by the continued thickening of a single shear band at a constant flow stress

Most of the plastic strain is accumulated during the sliding process, indicating that nucleation and initial propagation of a shear band are very brief. Analysis of a thermodynamic model also suggests a specific yield stress that is required to nucleate a shear band, after which the shear band is allowed to grow unconstrained.

In addition to the individual shear band, the STZ dynamics model has been adapted to investigate a collection of shear band events, which leads to the transition in flow serration in the inhomogeneous deformation regime. An interesting deformation phenomenon associated with metallic glass is that the shear band density and degree of flow serration are highly strain rate dependent, though the yield point of these materials is often independent of strain rate for rates up to $10^2 \sim 10^3$ s$^{-1}$ (Schuh et al. 2007, 2004). Low strain rates are characterized by strongly serrated flow, meaning that strain accumulates in the material in temporal bursts accompanied by relaxation stress drops resulting in a jagged stress-strain curve (Dalla Torre et al. 2010; Song et al. 2008). Higher strain rates are characterized by moderately serrated flow, and very high strain rates have little or no flow serration. The mechanisms that underlie the transition are hypothesized to be the competition between shear band nucleation and propagation (Schuh et al. 2004).

The STZ dynamics framework exhibits a yield point that is inherently rate dependent. To correct for this and make the yield point rate independent, the STZ dynamics model is extended to incorporate a strain-rate-dependent STZ volume and activation energy, given by the following log-linear forms:

$$\Omega_0 = -0.2\log_{10}\dot{\varepsilon} + 1.6 \quad \left[\text{nm}^3\right]$$
$$\Delta\text{F} = -0.12\log_{10}\dot{\varepsilon} + 1.07945 \quad [\text{eV}] \tag{13}$$

This parameterization of STZ volume and STZ energy barrier as a function of strain rate is not unique but aims to capture physical mechanisms of STZs (Dubach et al. 2009; Harris et al. 2016). The physical origin of the strain rate dependence
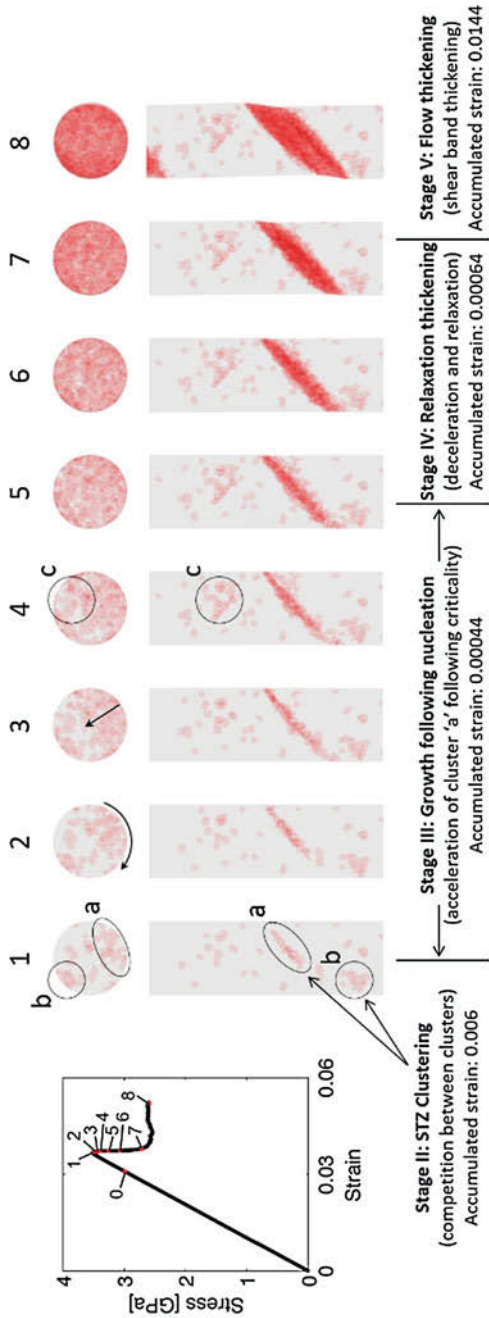
**Fig. 7** Shear localization process into one shear band. Numbered snapshots of the model metallic glass at various times during a constant strain rate, uniaxial tensile simulation. Each snapshot includes a semitransparent top and side view of the simulation cell. The evolution of the shear band is divided into four different stages following the initial elastic response. (Figure reproduced with the permission from Homer (2014))
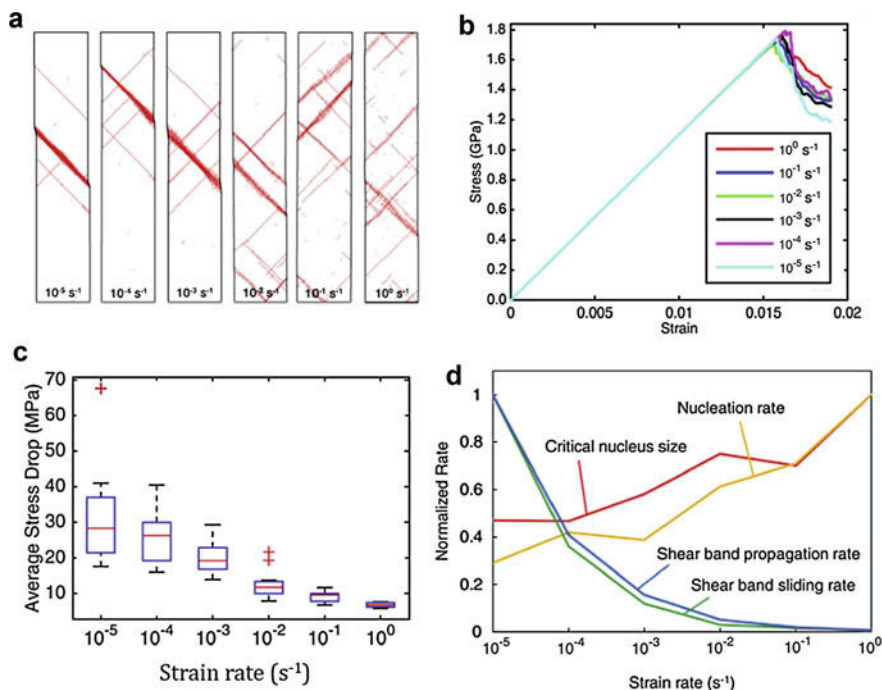
**Fig. 8** (**a**) Example simulations at various strain rates. Note the clear trend of increased shear band density with strain rate, and the increased appearance of free STZs at higher strain rates. (**b**) Stress-strain curves for the six simulations shown in (**a**). (**c**) Box plot of the average stress drop magnitude in each simulation, arranged by strain rate. (**d**) The comparison of median values of the normalized shear band nucleation rate (yellow), shear band propagation rate (blue), and shear band sliding rate (green). (Figure reproduced with permission from Harris et al. (2016))

remains an open question, possibly contributing to the time-dependent structural relaxation (Dubach et al. 2009) or a kinetic feature of the atomic motions associated with an STZ.

A transition of shear band density and morphology is captured by the model with increasing strain rate from $10^{-5}$ s$^{-1}$ to 1 s$^{-1}$. Figure 8a shows a group of six simulations, one from each strain rate studied, with increasing strain rate from left to right. Each simulation is at the final strain value of 1.9%. In general, low strain rates result in fewer, more dominant shear bands, with very few free STZs scattered outside the bands, while high strain rates feature larger numbers of less dominant shear bands, with many free STZs randomly scattered outside the bands. The corresponding stress-strain curves for these six simulations are displayed in Fig. 8b, showing a tightly grouped yield strength around 1.72 GPa. While the yield strengths are similar, low strain rates tend to relax more quickly after yield and have a lower flow stress than high strain rates. Examination of the flow serration regarding stress drop after yielding shows a negative correlation with strain rate (Fig. 8c). This

is in line with the experimental observations: strongly serrated flow at low strain rate and moderately serrated flow at higher strain rate.

Further study of shear band nucleation rates, propagation rates, and sliding rates in each simulation shows a transition from propagation-dominated shear banding at low strain rates to nucleation-dominated shear banding at high strain rates. A summary of the different statistics and their strain rate dependence is illustrated in Fig. 8d, where each rate has been scaled so they can be compared side by side. The underlying cause for the flow transition is hypothesized to be a strain-rate-dependent critical shear band nucleus size that increases with increasing strain rate (Harris et al. 2016). This is best illustrated by examining the stages of shear banding (ref. to Fig. 7). In the nucleation stage (II), STZs appear and begin to cluster into shear band nuclei, which grow and proliferate in the absence of a dominant shear band. If the strain rate is low, then a small critical nucleus size means that the growth stage (III) is reached quickly, and one shear band rapidly propagates across the sample and begins to dominate all plasticity in the sample. Then, in the sliding stage (IV and V), additional plasticity is concentrated in bursts on that dominant band. If instead the strain rate is high, then a large critical nucleus size means that the growth stage (III) is delayed, or skipped entirely, and plasticity continues to be accommodated by nucleation of additional shear band nuclei in the nucleation stage (II). Then sliding stage (IV and V) occurs more gradually as shear band nuclei begin to intersect each other, and plasticity remains relatively diffuse.

In essence, the STZ dynamics model provides insight into the shear banding process, contributing toward a better picture incorporating both kinetic and thermodynamic nucleation criteria of shear band formation. These types of studies demonstrate the strength of mesoscale models in elucidating the micromechanics behind the macroscopic process. Some of these features would be difficult to observe by other techniques.

## 3.3   Nanoscale Strengthening Subjected to Cyclic Nanoindentation

Metallic glasses exhibit a broad range of interesting phenomena due to the inherently complex non-equilibrium states, one of which is that they can exhibit nanoscale strengthening subjected to cyclic nanoindentation in the elastic regime. This has been demonstrated in nanoindentation experiment, showing a statistical increase in strength as a result of cyclic loading at a magnitude before the first significant plastic event (e.g., shear band), signified by a pop-in on the load-displacement curve (Packard et al. 2010, 2008). Interestingly, the cyclic strengthening can only occur if the cycling is of a sufficient magnitude, if the indenter is actually cycled (holding a constant load of equal magnitude and time does not lead to strengthening), and the strengthening saturates after a finite amount of cycles.
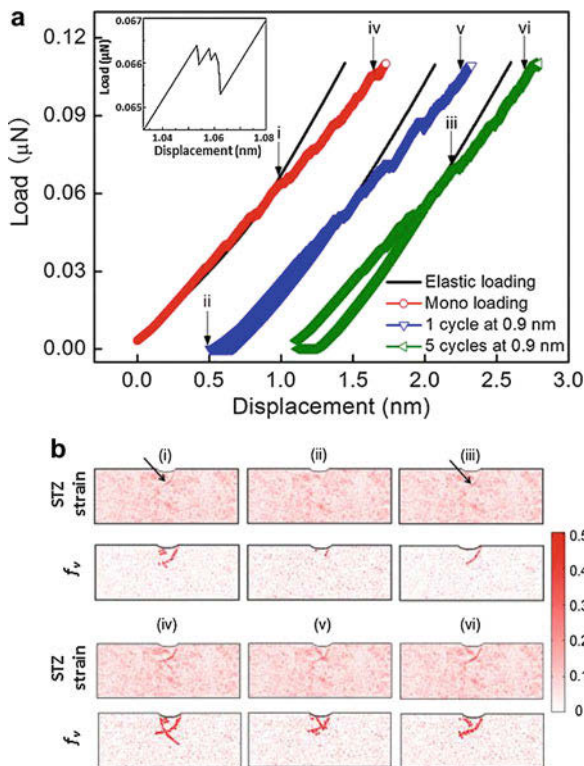
**Fig. 9** Simulated nanoindentation results for various loading conditions. (**a**) Load-displacement curves for the monotonic loading, one-cycle and five-cycle loading with the cyclic depth of 0.9 nm. Vertical arrows indicate the points: (i) at the yield of the monotonic loading; (ii) at the reload after the first cycle; (iii) at yield after the five-cycle loading; and (iv), (v), and (vi) at a post-yield load of 0.1 μN for the three loading conditions. The origin of the one-cycle and five-cycle loading curves are shifted for a clear view. The inset shows an enlarged view of load drops around (i). (**b**) The snapshots display the spatial distribution of STZ strain and excess free volume $f_v$ at points (i–vi). Black arrows denote the STZ strain accumulation at (i) and (iii). (Figure reproduced with permission from Wang et al. (2015))

To elucidate the underlying mechanisms that cause the strengthening, the STZ dynamics framework is adapted by including contact mechanics in the finite element analysis solver (Packard et al. 2011, 2010). Furthermore, the model incorporates excess free volume as a state variable, to study the interplay of glass deformation and structural evolution under cyclic indentation tests at an experimentally relevant time scale (Wang et al. 2015).

Results from simulations under various loading conditions are illustrated in Fig. 9. For a clear view, the origins of the load-displacement curves with cyclic loading are shifted to the right. In Fig. 9a, from left to right, the three curves represent the monotonic loading, one-cycle and five-cycle loadings with the cyclic depth of 0.9 nm, respectively. The yield point of each test is indicated by an

arrow, identified by an applicable deviation from the elastic response. The yield load increases with the number of cycles, and after five cycles, it is ∼ 10% higher than that of the monotonic response, which is consistent with the nanoscale cyclic strengthening observed experimentally. Additionally, the snapshots of the spatial distributions of STZ strain and excess free volume at several critical moments are displayed in Fig. 9b. The accumulation of excess free volume can be clearly observed at (i) and (iii), indicating extensive STZ activity prior to the yield point. The cyclic loading can lead to structural change reflected by the annihilation of excess free volume. In a comparison of excess free volume distribution at (i) with (iii), the cluster on the top left in (i) almost disappears after five cycles in (iii). This decrease of excess free volume after cyclic loading gives rise to the mechanical strengthening as a consequence of removal of mechanically weak sites.

The nanoindentation simulation has further been used to detect the cause of nanoscale strengthening subject to cyclic loading. For instance, Fig. 10a displays the cumulative distribution of the yield loads at ten different indentation locations for monotonic and cyclic loadings after one, three, and five cycles, at the cyclic depths of 0.9 nm. Notably, there exists a large distribution of yield loads for monotonic loading, e.g., the minimum yield load is about half of the maximum one. Further, as the number of cycles increases, the distribution curves become sharper and shift to the right particularly at the lower tail, indicating that the "weak" samples are strengthened during cyclic loading. Such effect could be contributed to the removal of the residual stress (Wang et al. 2011a), structural relaxation (Pan et al. 2009), and the arrest of the shear band (Yang et al. 2006). With excess free volume as a state variable, the STZ dynamics simulations demonstrate that the strengthening is directly related to the decrease of large excess free volume sites. When the loading cycles increase from 0 (i.e., monotonic) to 5, the excess free volume is reduced progressively, signified by a left shift of the distribution curves shown in Fig. 10b.
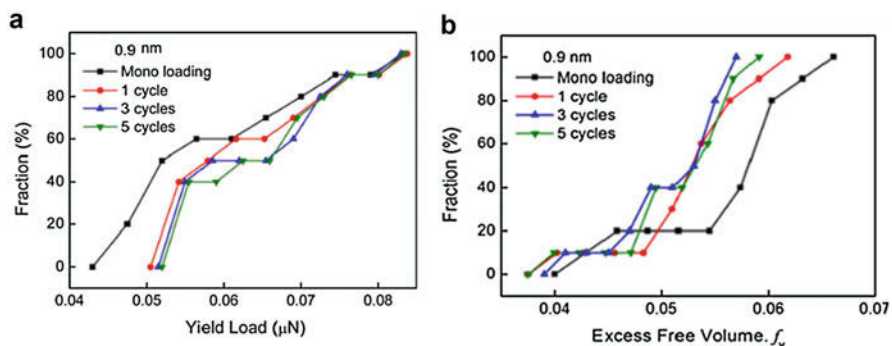


**Fig. 10** (**a**) Cumulative distributions of the yield loads for monotonic loading, and various cyclic loadings after one, three, and five cycles at the cyclic depth of 0.9 nm. (**b**) Cumulative distribution curves of the excess free volume $f_v$ in a selected region for monotonic loading and various cyclic loadings after one, three, and five cycles at the depth of 0.9 nm. (Figure adapted with permission from Wang et al. (2015))

The reduction of excess free volume has been ascribed to the observed cyclic strengthening, since it results in a reduced rate of microplastic events by increasing STZ activation energy. Additionally, the left shift becomes less pronounced after one cycle and saturates after five cycles, which is consistent with the saturation of the strengthening effect. Overall, the STZ dynamics simulations indicate that the energetics and timescales of STZ activity are plausible as a mechanism to cause structural evolution that is consistent with nanoscale strengthening.

The application of a mesoscale technique to investigate the nanomechanics of experiments indicates the strong potential to elucidate phenomena that are difficult to measure by experimental techniques.

## 4    Conclusions and Outlooks

The STZ dynamics model, combining a coarse-grained approach and the kMC algorithm, provides a useful framework to investigate the deformation behaviors of metallic glasses. On the one hand, the coarse-graining enables collections of atoms, or STZs, to be tracked; on the other hand, the kMC algorithm allows the stochastic activation of these STZs based on the energetics of the system. The combination of these two features enables the simulation of deformation behavior at larger time and length scales while preserving a microscopic view of the processes that dominate deformation.

The STZ dynamics model has been used to investigate the deformation behaviors of metallic glasses in a variety of conditions. The modeling technique captures the overall deformation behaviors expected of metallic glasses and the underlying spatial and temporal correlations of STZs that contribute to different deformation modes. The mesoscale model provides details into the formation of individual shear band and the collection of shear band events which leads to the transition of flow serration in the inhomogeneous deformation regime. Insights into nanoindentation experiments are possible through the contact mechanics adaptions. The interplay of deformation and structural evolution is accessed via the incorporation of a free volume state variable.

The STZ dynamics framework will continue to be useful in the investigation of the mechanical behaviors of metallic glasses. Many challenges remain for further development, which includes, among others:

- *Activation energy functional:* The development and implementation of activation energy functionals that more accurately capture the nature and variability of disordered glassy structure and the structural dynamics of metallic glasses.
- *Failure mechanisms:* The failure of metallic glasses involves the strain softening, adiabatic heating, cavitation, and crack formation in shear bands. One could incorporate heat and mass transfer constitutive relations that are capture conditions leading up to failure. One could also include an additional stochastic process for cavitation that precedes crack formation. Finally, once a crack is

initiated, one could model the crack propagation using standard finite element techniques.

- *Adaptive coarse-graining:* Since the scales of STZs and shear bands are different, one could use adaptive meshing to optimize the simulations. Depending on the nature of what is happening, one could coarsen or refine a mesh. For example, after a shear band appears, one could deal with groups of STZs instead of dealing with individual STZs. Mesh refinement/remeshing may be geometrically challenging but can be handled in practice with the use of advanced adaptive meshing algorithms available in finite element packages. Mesh coarsening could also be handled but would require the knowledge of the dynamic evolution of STZ activation volume.

- *Large-scale simulations:* While powerful, current implementations of the STZ dynamics framework are limited in their ability to simulate large structure. A parallel, distributed memory implementation of the STZ dynamics modeling would significantly increase the size of systems that can be examined by the technique. In particular, a different implementation might allow the use of high-performance computing systems. Large-scale simulations would help answer questions about shear banding phenomena beyond the initial stage of shear band formation, which is essential to directly connect with metallic glass toughness and failure at the macroscopic scale.

# References

Albaret T, Tanguy A, Boioli F, Rodney D (2016) Mapping between atomistic simulations and Eshelby inclusions in the shear deformation of an amorphous silicon model. Phys Rev E 93:053002

Argon AS (1979) Plastic deformation in metallic glasses. Acta Metall 27:47–58

Baret J-C, Vandembroucq D, Roux S (2002) Extremal model for amorphous media plasticity. Phys Rev Lett 89:195506

Boioli F, Albaret T, Rodney D (2017) Shear transformation distribution and activation in glasses at the atomic scale. Phys Rev E 95:033005

Bulatov VV, Argon AS (1994a) A stochastic model for continuum elasto-plastic behavior. I. Numerical approach and strain localization. Model Simul Mater Sci Eng 2:167

Bulatov VV, Argon AS (1994b) A stochastic model for continuum elasto-plastic behavior. II. A study of the glass transition and structural relaxation. Model Simul Mater Sci Eng 2:185

Bulatov VV, Argon AS (1994c) A stochastic model for continuum elasto-plastic behavior. III. Plasticity in ordered versus disordered solids. Model Simul Mater Sci Eng 2:203

Cheng YQ, Ma E (2011) Atomic-level structure and structure–property relationship in metallic glasses. Prog Mater Sci 56:379–473

Dahmen KA, Ben-Zion Y, Uhl JT (2009) Micromechanical model for deformation in solids with universal predictions for stress-strain curves and slip avalanches. Phys Rev Lett 102:175501

Dalla Torre FH, Klaumünzer D, Maaß R, Löffler JF (2010) Stick–slip behavior of serrated flow during inhomogeneous deformation of bulk metallic glasses. Acta Mater 58:3742–3750

Ding J, Patinet S, Falk ML, Cheng Y, Ma E (2014) Soft spots and their structural signature in a metallic glass. Proc Natl Acad Sci 111:14052–14056

Dubach A, Dalla Torre FH, Löffler JF (2009) Constitutive model for inhomogeneous flow in bulk metallic glasses. Acta Mater 57:881–892

Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. Proc R Soc Lond A Math Phys Sci 241:376

Falk ML, Maloney CE (2010) Simulating the mechanical response of amorphous solids using atomistic methods. Eur Phys J B 75:405–413

Fan Y, Iwashita T, Egami T (2014) How thermally activated deformation starts in metallic glass. Nat Commun 5:5083

Fan Y, Iwashita T, Egami T (2017) Energy landscape-driven non-equilibrium evolution of inherent structure in disordered material. Nat Commun 8:15417

Fusco C, Albaret T, Tanguy A (2010) Role of local order in the small-scale plasticity of model amorphous materials. Phys Rev E 82:066116

Greer AL, Cheng YQ, Ma E (2013) Shear bands in metallic glasses. Mater Sci Eng 74:71–132

Hardin TJ, Homer ER (2015) Microstructural factors of strain delocalization in model metallic glass matrix composites. Acta Mater 83:203–215

Harris MB, Watts LS, Homer ER (2016) Competition between shear band nucleation and propagation across rate-dependent flow transitions in a model metallic glass. Acta Mater 111:273–282

Homer ER (2014) Examining the initial stages of shear localization in amorphous metals. Acta Mater 63:44–53

Homer ER, Schuh CA (2009) Mesoscale modeling of amorphous metals by shear transformation zone dynamics. Acta Mater 57:2823–2833

Homer ER, Schuh CA (2010) Three-dimensional shear transformation zone dynamics model for amorphous metals. Model Simul Mater Sci Eng 18:065009

Homer ER, Rodney D, Schuh CA (2010) Kinetic Monte Carlo study of activated states and correlated shear-transformation-zone activity during the deformation of an amorphous metal. Phys Rev B 81:064204

Hufnagel TC, Schuh CA, Falk ML (2016) Deformation of metallic glasses: recent developments in theory, simulations, and experiments. Acta Mater 109:375–393

Hwang J, Melgarejo ZH, Kalay YE, Kalay I, Kramer MJ, Stone DS, Voyles PM (2012) Nanoscale structure and structural relaxation in $Zr_{50}Cu_{45}Al_5$ Bulk Metallic Glass. Phys Rev Lett 108:195505

Johnson WL, Samwer K (2005) A universal criterion for plastic yielding of metallic glasses with a $(T/T_g)^{2/3}$ temperature dependence. Phys Rev Lett 95:195501

Li L, Homer ER, Schuh CA (2013) Shear transformation zone dynamics model for metallic glasses incorporating free volume as a state variable. Acta Mater 61:3347–3359

Li L, Wang N, Yan F (2014) Transient response in metallic glass deformation: a study based on shear transformation zone dynamics simulations. Scr Mater 80:25–28

Liu YH et al (2011) Characterization of nanoscale mechanical heterogeneity in a metallic glass by dynamic force microscopy. Phys Rev Lett 106:125504

Ma E, Ding J (2016) Tailoring structural inhomogeneities in metallic glasses to enable tensile ductility at room temperature. Mater Today 19:568–579

Ma E, Zhang Z (2010) Reflections from the glass maze. Nat Mater 10:10

Malek R, Mousseau N (2000) Dynamics of Lennard-Jones clusters: a characterization of the activation-relaxation technique. Phys Rev E 62:7723–7728

Maloney C, Lemaître A (2004) Universal breakdown of elasticity at the onset of material failure. Phys Rev Lett 93:195501

Maloney CE, Lemaître A (2006) Amorphous systems in athermal, quasistatic shear. Phys Rev E 74:016118

Packard CE, Witmer LM, Schuh CA (2008) Hardening of a metallic glass during cyclic loading in the elastic range. Appl Phys Lett 92:171911

Packard CE, Homer ER, Al-Aqeeli N, Schuh CA (2010) Cyclic hardening of metallic glasses under Hertzian contacts: Experiments and STZ dynamics simulations. Philos Mag 90: 1373–1390

Packard CE, Franke O, Homer ER, Schuh CA (2011) Nanoscale strength distribution in amorphous versus crystalline metals. J Mater Res 25:2251–2263

Pan D, Yokoyama Y, Fujita T, Liu YH, Kohara S, Inoue A, Chen MW (2009) Correlation between structural relaxation and shear transformation zone volume of a bulk metallic glass. Appl Phys Lett 95:141909

Picard G, Ajdari A, Bocquet L, Lequeux F (2002) Simple model for heterogeneous flows of yield stress fluids. Phys Rev E 66:051501

Rodney D, Schuh C (2009) Distribution of thermally activated plastic events in a flowing glass. Phys Rev Lett 102:235503

Rodney D, Tanguy A, Vandembroucq D (2011) Modeling the mechanics of amorphous solids at different length scale and time scale. Model Simul Mater Sci Eng 19:083001

Schuh CA, Lund AC, Nieh TG (2004) New regime of homogeneous flow in the deformation map of metallic glasses: elevated temperature nanoindentation experiments and mechanistic modeling. Acta Mater 52:5879–5891

Schuh CA, Hufnagel TC, Ramamurty U (2007) Mechanical behavior of amorphous alloys. Acta Mater 55:4067–4109

Song SX, Bei H, Wadsworth J, Nieh TG (2008) Flow serration in a Zr-based bulk metallic glass in compression at low strain rates. Intermetallics 16:813–818

Şopu D, Stukowski A, Stoica M, Scudino S (2017) Atomic-level processes of shear band nucleation in metallic glasses. Phys Rev Lett 119:195503

Spaepen F (1977) A microscopic mechanism for steady state inhomogeneous flow in metallic glasses. Acta Metall 25:407–415

Srolovitz D, Vitek V, Egami T (1983) An atomistic study of deformation of amorphous metals. Acta Metall 31:335–352

Su C, Anand L (2006) Plane strain indentation of a Zr-based metallic glass: experiments and numerical simulation. Acta Mater 54:179–189

Tsai P, Kranjc K, Flores KM (2017) Hierarchical heterogeneity and an elastic microstructure observed in a metallic glass alloy. Acta Mater 139:11–20

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Sickafus KE, Kotomin EA, Uberuaga BP (eds) Radiation effects in solids. Springer, Dordrecht, pp 1–23

Wang L, Bei H, Gao YF, Lu ZP, Nieh TG (2011a) Effect of residual stresses on the hardness of bulk metallic glasses. Acta Mater 59:2858–2864

Wang N, Ding J, Yan F, Asta M, Ritchie RO, Li L (2018) Spatial correlation of elastic heterogeneity tunes the deformation behavior of metallic glasses. npj Comput Mater 4:19

Wang N, Yan F, Li L (2015) Mesoscopic examination of cyclic hardening in metallic glass. J Non-Cryst Solids 428:146–150

Wang WH, Wen P, Zhao DQ, Pan MX, Wang RJ (2011b) Relationship between glass transition temperature and Debye temperature in bulk metallic glasses. J Mater Res 18:2747–2751

Xu B, Falk M, Li J, Kong L (2017) Strain-dependent activation energy of shear transformation in metallic glasses. Phys Rev B 95:144201

Yang B, Riester L, Nieh TG (2006) Strain hardening and recovery in a bulk metallic glass under nanoindentation. Scr Mater 54:1277–1280

Zhao P, Li J, Wang Y (2013) Heterogeneously randomized STZ model of metallic glasses: softening and extreme value statistics during deformation. Int J Plast 40:1–22

Zink M, Samwer K, Johnson WL, Mayr SG (2006) Plastic deformation of metallic glasses: size of shear transformation zones from molecular dynamics simulations. Phys Rev B 73:172203

# Kinetic Monte Carlo Modeling of Martensitic Phase Transformation Dynamics

# 57

Ying Chen

## Contents

### Abstract

Martensitic transformation is, on the one hand, a pervasive deformation mechanism in both structural and functional materials, and on the other hand, a first-order phase transition that is diffusionless. The transformation from one crystal structure to another takes place by a volumetric change and a large shear. As a result, modeling of the martensitic transformation process requires an integration of thermodynamic, kinetic, and mechanical considerations. Moreover, the transformation process is intrinsically stochastic. Not only is there a competition between nucleation and propagation modes, but there are also

Y. Chen (✉)
Department of Materials Science and Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
e-mail: cheny20@rpi.edu

competitions among different ways of transformation and different regions for transformation. In this chapter, an integrated thermodynamic and Kinetic Monte Carlo (KMC) treatment of martensitic transformation is presented. Modeling martensitic transformation as a unit process, the free energy function for potential transformation of each unit is determined. Stress and strain distributions are predicted by the Finite Element method after each unit transformation and are incorporated in the free energy function. A KMC algorithm that incorporates the free energy function in the rate formula is invoked to select a unit to transform and advance the time. The modeling formulation is described in detail, so is the simulation algorithm. Examples of transformation dynamics modeled by this method will be shown.

## 1    Introduction

Martensitic transformation is a solid-state, diffusionless, and displacive phase change that typically involves conversion between different crystal structures (Otsuka et al. 1979; Wechsler et al. 1953). It is common that the austenite phase (which is sometimes also referred to as the parent phase) has a higher symmetry (e.g., the most common symmetry for austenite is cubic, such as B2, D0$_3$, and L2$_1$ crystal structures), while the martensitic phase has a lower symmetry. Reversible martensitic transformations are commonly observed in Shape Memory Alloys (SMAs) such as Ti-Ni, Cu-Al-Ni, Ni-Mn-Ga, Co-Ni-Al, and Fe-Pd (Lagoudas 2008). SMAs can respond promptly and repeatedly to thermal, mechanical, and combined stimuli and switch between two different shapes under thermal or mechanical cycling (Karaca et al. 2007; Otsuka and Wayman 1998; Tadaki et al. 1988). Many ceramics also undergo martensitic transformation, such as tetragonal to monoclinic in zirconia, cubic to tetragonal in lead titanate, and cubic to tetragonal in cristobalite (Kriven 1995). Martensitic transformation is also an important deformation mechanism in many structural metals, such as steels (Tao et al. 2007) and titanium alloys (Jaworski and Ankem 2005). This chapter mainly discusses modeling of reversible martensitic phase transformations in shape memory alloys and ceramics, which can be adapted for irreversible transformation.

Computational models for reversible martensitic transformations are relatively rare. There are several molecular dynamics studies (Guda Vishnu and Strachan 2012; Kastner et al. 2011; Tatar and Kazanc 2012; Zhong et al. 2012), but the length (e.g., 10-30 nm) and time (e.g., nanoseconds) scales accessed are not directly comparable to those (e.g., hundreds of nanometers (Ozdemir et al. 2012) to micrometers in size and milliseconds (Juan et al. 2009) to minutes in time) in most experiments. There are also a number of finite element (Gall et al. 2000; Gall and Sehitoglu 1999; Lagoudas et al. 2012; Manchiraju and Anderson 2010; Manchiraju et al. 2011; Patoor et al. 1995; Wood and Clyne 2006) and phase field (Jin et al. 2001; Levitas et al. 2009, 2013; Mamivand et al. 2013, 2014) models, which can incorporate larger scales but generally tend to lack connections between simulated scales and those intrinsic to transformation. Moreover, the inherent

variability in martensite nucleation and growth and dependence of such variability on microstructure have rarely been taken into consideration in these prior models.

Kinetic Monte Carlo (KMC) is a stochastic modeling method widely used to model microstructure evolution. It has also been used to simulate deformation in amorphous solids which is treated as a stochastic sequence of local inelastic distortions in a lattice model (Bulatov and Argon 1994a, b, c) or a continuum model (Homer et al. 2010; Homer and Schuh 2009). Compared to shear-dominant deformation in amorphous solids mentioned above, martensitic transformation is subject to crystallographic and thermodynamic constraints. In this chapter, a KMC model that treats martensitic transformations as a sequence of stochastic unit processes (Chen and Schuh 2015) is described.

## 2  Model Overview

The stochastic physics-based modeling framework for reversible martensitic transformation (Chen and Schuh 2015) integrates thermodynamic description of transformation, kinetic Monte Carlo (KMC) treatment of phase evolution, and Finite Element Method (FEM) modeling of stress and strain distribution. A thermodynamic energy landscape description of austenitic and martensitic states which governs the transformations is examined in Sect. 3. Section 4 describes how the onset and progression of the transformations are controlled by KMC using a transition-state rate equation that incorporates local thermomechanical state. KMC modeling includes a physical timescale for transformation, enabling the exploration of stochastic transformation dynamics. In Sect. 5, FEM modeling procedures are provided. FEM is utilized to apply the displacive transformations and predict mechanical state everywhere in the material, which allows the local state to affect the transformation sequence. Section 6 discusses some considerations for computational implementation of the model and shows some examples.

The model offers the following unique capabilities: (i) capturing transformation stochasticity, (ii) predicting the evolution in spatial distributions of phases and stress/strain, (iii) predicting transformation process based on energetics incorporating local thermomechanical states, (iv) using a physical time scale and a microstructure length scale, and (v) predicting two-way (i.e., reversible) transformations between austenite and martensite phases.

## 3  Thermodynamic Energy Landscape for Martensitic Transformation

In the following, subscripts of "A" and "M" are used to denote properties of austenite and martensite, respectively. The Gibbs free energies $G_A$ and $G_M$ for austenite and martensite per unit volume as a function of temperature $T$ are illustrated in Fig. 1a. When there is no external stress, $G_A = G_A^{ch} = H_A - T S_A$

**(a) Energy-Temperature Diagram**



**(b) Stress-Temperature Diagram**

**(c) Energy Landscape for Transformation States**
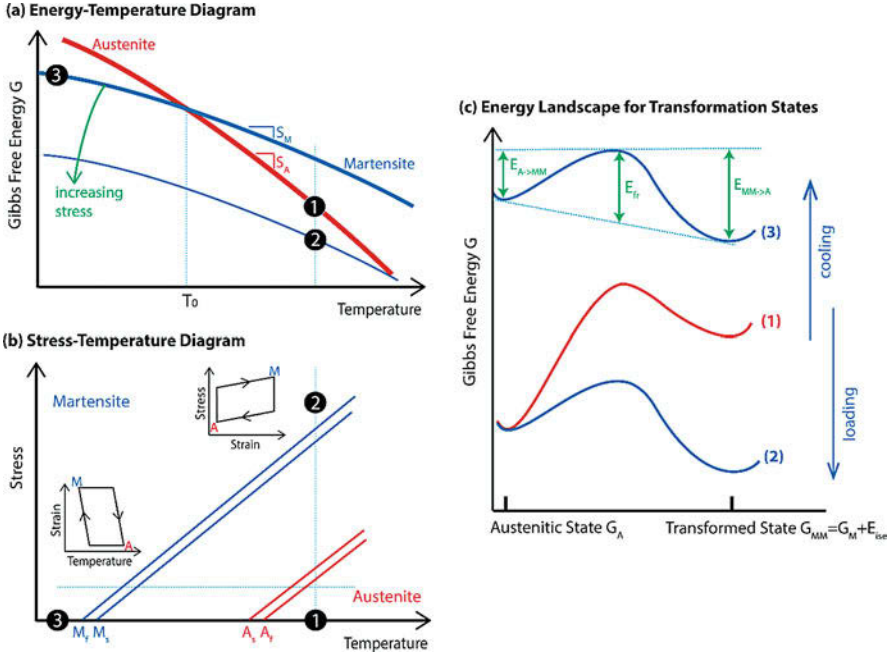
**Fig. 1** (**a**) The Gibbs free energy for the austenite phase, $G_A$, (in red) and the free energy for the martensite phase, $G_M$, (in blue) as a function of the temperature $T$. $T_0$ denotes the equilibrium temperature at which $G_A = G_M$. $G_M$ decreases with increasing resolved shear stress. (**b**) The stress-temperature phase diagram for martensite and austenite. The insets show strain excursions as a result of thermal cycling or mechanical cycling. (**c**) The energy landscape between the austenitic state and the transformed martensitic state for three different conditions which have been marked in both (a) and (b). Condition 1 corresponds to a high temperature and zero-stress condition; loading from this condition leads to Condition 2 while cooling from this condition leads to Condition 3

(red line) and $G_M = G_M^{ch} = H_M - T S_M$ (bold blue line), where $G_A^{ch}$ and $G_M^{ch}$ are the chemical free energies for austenite and martensite, respectively. The slopes of these lines are the entropy $S_A$ and $S_M$, respectively; $S_A > S_M$ and accordingly $G_A^{ch}$ decreases with $T$ more rapidly than $G_M^{ch}$. Let $T_0$ denote the equilibrium temperature at which $G_{A,0}^{ch} = G_{M,0}^{ch}$ and the transformation enthalpy $\Delta H_0 = H_{M,0} - H_{A,0} = T_0(S_{M,0} - S_{A,0}) = T_0 \Delta S_0$, where $\Delta H_0 < 0$ and the transformation entropy $\Delta S_0 = S_{M,0} - S_{A,0} < 0$. If the change in heat capacity is assumed insignificant, $\Delta H \approx \Delta H_0$ and $\Delta S \approx \Delta S_0$. The difference in chemical free energy at temperature $T$ can be expressed as

$$\Delta G^{ch} = G_M^{ch} - G_A^{ch} = -\Delta S_0 (T - T_0) \tag{1}$$

when $T < T_0$, $\Delta G^{ch} < 0$ and martensite is the thermodynamically preferred phase (e.g., state 3 marked on the blue curve in Fig. 1a); when $T > T_0$, $\Delta G^{ch} > 0$ and

austenite is the stable phase (e.g., state 1 marked on the red curve in Fig. 1a). Upon application of an external stress that leads to a resolved shear stress $\tau$ on the habit plane (in the elastic regime before yielding), we estimate $G_A \approx G_A^{ch}$ and $G_M \approx G_M^{ch} - \tau\gamma_0$ because the elastic strain is insignificant compared to the transformation simple shear strain $\gamma_0$. $G_M$ decreases with increasing stress, shown by the thinner blue line in Fig. 1a.

$$\Delta G = G_M - G_A = \Delta G^{ch} - \tau\gamma_0 = -\Delta S_0 \left(T - T_0\right) - \tau\gamma_0 \qquad (2)$$

It is now possible to have $\Delta G < 0$ and accordingly a stable martensitic state even when $T > T_0$ so long as $\tau$ is sufficiently high. For example, at the same temperature, the stable phase changes from austenite (marked as state 1 in Fig. 1a) to martensite (marked as state 2 in Fig. 1a) under finite $\tau$. A third term, $P\varepsilon_V$, where $P$ is the pressure, may be introduced into Eq. (2) for transformations involving a significant volumetric strain $\varepsilon_V$.

The effect of temperature and stress is further depicted in the stress-temperature diagram in Fig. 1b, where the three states marked in Fig. 1a are also highlighted. At low temperatures and high stresses, martensite is more stable, while at high temperatures and low stresses austenite is the preferred phase. Conversions between the two phases can be achieved by many ways of thermomechanical excursions, such as thermal cycling under no stress, thermal cycling at a finite stress, and isothermal mechanical cycling. When there is no stress, transformation from austenite to martensite under cooling is characterized by a martensitic start temperature $M_s$ and a martensitic finish temperature $M_f$, and reverse transformation from martensite to austenite upon heating starts at the austenite start temperature $A_s$ and completes at the austenite finish temperature $A_f$. At a constant stress, phase conversion induced by thermal cycling is accompanied by a recoverable strain as illustrated in the bottom inset in Fig. 1b, leading to the so-called two-way shape memory effect. Similarly, mechanical cycling at a constant temperature leads to a phenomenon called superelasticity, which describes the recoverability of stress-induced strain (see the top inset in Fig. 1b) due to the reversibility of the phase conversion.

Martensitic transformation involves energy penalties, rendering the free energy of the transformed state, $G_{MM}$, different from that of the martensite phase, $G_M$. The difference between them corresponds to the recoverable part of the energy penalty, which is defined as $E_{ise}$, whose subscripts "ise" stand for contributions from interfacial energy, surface energy, and elastic strain energies.

$$E_{ise} = \gamma_i A_i + \Delta\gamma_{sf} A_{sf} + E_{el} \qquad (3)$$

where $\gamma_i$ is the austenite/martensite interface energy per unit area and $A_i$ is the change in interfacial area density. For an austenite domain surrounded by $n_A$ and $n_M$ austenite and martensite nearest neighbors, respectively, $A_i = (n_A - n_M)A_b$, where $A_b$ is the area density of each domain boundary. $\Delta\gamma_{sf} = \gamma_{sf}^M - \gamma_{sf}^A$ is the difference in surface energy per unit area, and multiplies the specific surface area $A_{sf}$; this term

pertains to the intersection of the phases with free surfaces and is usually considered negligibly small (Chen and Schuh 2011). $E_{el}$ is the increase in elastic strain energy density. Let's define

$$G_{MM} = G_M + E_{ise} \approx G_M^{ch} - \tau\gamma_0 + E_{ise} \qquad (4)$$

as the free energy of a transformed martensitic state. During transformation from the initial austenitic state with free energy $G_A$ to the transformed state with free energy $G_{MM}$, there is energy dissipation mainly resulting from the frictional work $E_{fr}$, which is irreversible and is positive in both transformation directions (Bonnot et al. 2008; Ortín and Planes 1988; Wollants et al. 1993).

Figure 1c shows the energy landscape for transformation between the initial and the transformed states, with an arbitrary reaction coordinate along the x-axis (which may be viewed as transformation strain). It is constructed by imposing a kinetic barrier with height equal to $E_{fr}$ above the average of $G_A$ and $G_{MM}$ (as demonstrated for curve #3 in Fig. 1c). With the energy maxima corresponding to the activated state being $(G_{MM} + G_A)/2 + E_{fr}$, and further utilizing Eqs. (2) and (4), the activation energy barriers for transformations between the two states can be written as

$$\begin{cases} E_{A\to MM} = \left[\frac{1}{2}(G_{MM} + G_A) + E_{fr}\right] - G_A = -\frac{1}{2}\Delta S_0(T - T_0) - \frac{1}{2}\tau\gamma_0 + \frac{1}{2}E_{ise} + E_{fr} \\ E_{MM\to A} = \left[\frac{1}{2}(G_{MM} + G_A) + E_{fr}\right] - G_{MM} = \frac{1}{2}\Delta S_0(T - T_0) + \frac{1}{2}\tau\gamma_0 - \frac{1}{2}E_{ise} + E_{fr} \end{cases}$$
$$(5)$$

where $E_{fr}$ may be treated as an intrinsic frictional resistance to transformation (i.e., a constant), allowing $E_{A\to MM}$ and $E_{MM\to A}$ to be determined at a given time. $E_{fr}^{A\to MM} = E_{fr}^{MM\to A} = E_{fr}$ is assumed for simplicity. The energy barriers for forward transformation (A → MM) and reverse transformation (MM → A) have opposite trends with respect to changes in the temperature $T$ and the stress $\tau$.

The three curves in Fig. 1c are for the same three conditions highlighted in Fig. 1a, b. Consider condition 1 (high temperature and zero stress) as the initial condition (red curve). $G_A$ is much lower than $G_{MM}$, and the stable state is the austenitic state. With decreasing temperature (from condition 1 to condition 3), both $G_A$ and $G_{MM}$ increase, but $G_A$ increases more rapidly than $G_{MM}$. With sufficient cooling, $G_A$ becomes higher than $G_{MM}$, and the transformed state becomes the preferred state. During cooling, $E_{A\to MM}$ decreases and $E_{MM\to A}$ increases as $\Delta S_0 < 0$ in Eq. (5). When $E_{A\to MM}$ becomes sufficiently low, thermal activation may overcome the barrier enabling transformation from the initial austenitic state to the martensitic state. On the other hand, upon loading from condition 1 to condition 2, $G_A$ remains nearly the same as the temperature is not changed while $G_{MM}$ decreases rapidly and becomes lower than $G_A$ when the applied load is sufficiently high. A high resolved shear stress $\tau$ also leads to a decrease in the energy barrier $E_{A\to MM}$ so that transformation to the martensitic state is not only thermodynamically preferred but also kinetically possible. On both curve #2 and curve #3 in Fig. 1c, the energy barrier for reverse transformation, $E_{MM\to A}$, is very high, suggesting that reverse transformation is possible but with a very low probability.

In Eqs. (2), (4), and (5), the resolved shear stress $\tau$ depends on the specific habit plane and shear direction and can be different for different martensite variants. For $i$th and $j$th variants, Eq. (2) leads to $G_{M_j} - G_{M_i} = (\tau_i - \tau_j)\,\gamma_0$. The energy barrier for $MM_i \to MM_j$ conversion may be determined in a similar manner to Eq. (5):

$$E_{MM_i \to MM_j} = \frac{1}{2}(\tau_i - \tau_j)\,\gamma_0 + E'_{fr} \tag{6}$$

where $E'_{fr}$ is the frictional work associated with inter-variant conversion. By the same token, extension to polycrystalline systems requires identification of transformation shear planes and directions in each grain.

## 4        KMC Modeling of Transformation Progression

KMC is an iterative modeling method. The modeling steps in each simulation step/iteration are provided in the flow chart in Fig. 2.

In each modeling iteration, the activation rates (in 1/s) for transformations between austenite and the $i$th martensite variant state are obtained by multiplying an attempt frequency $\nu_0$ with the activation probability.

$$\begin{cases} \omega_{A \to MM_i} = \nu_0 \exp\left(-\frac{E_{A \to MM_i}\Omega_0}{kT}\right) \\ \omega_{MM_i \to A} = \nu_0 \exp\left(-\frac{E_{MM_i \to A}\Omega_0}{kT}\right) \end{cases} \tag{7}$$

where $k$ is Boltzmann's constant and $\Omega_0$ can be considered a fundamental domain volume for a "unit" of transformation. The activation rate for inter-variant state conversion can be calculated similarly.

$$\omega_{MM_i \to MM_j} = \nu_0 \exp\left(-\frac{E_{MM_i \to MM_j}\Omega_0}{kT}\right) \tag{8}$$

The cumulative activation rate $\omega$ is the summation of individual activation rates of the $N_A$ austenite domains and $N_M$ martensite domains in the model system containing $N_A + N_M$ domains.

$$\omega = \sum_{n=1}^{N_A}\left(\sum_{i=1}^{N_V}\omega_{A \to MM_i}^n\right) + \sum_{m=1}^{N_M}\left(\omega_{MM_i \to A}^m + \sum_{j=1}^{N_V-1}\omega_{MM_i \to MM_j}^m\right) \tag{9}$$

where $N_V$ is the number of martensite variants. An austenite domain may transform to any of the $N_V$ martensite variants, while a given martensite variant may transform to either austenite or any of the other $N_V$ -1 variants. Forward and reverse transformations are treated "equally," i.e., they are both permitted at any time.
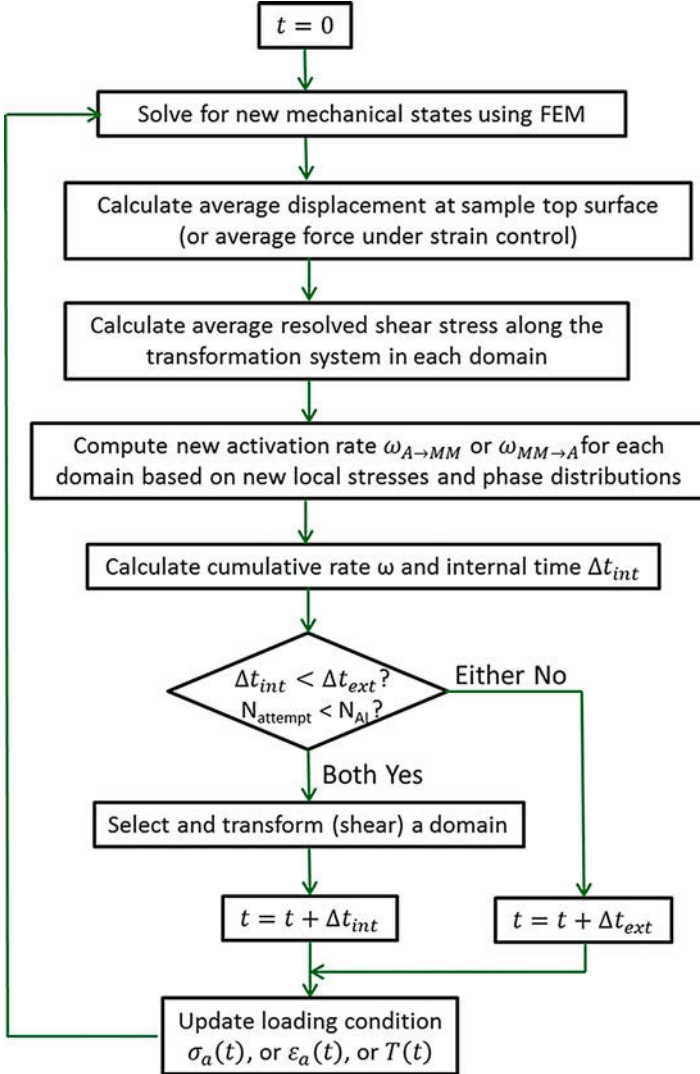
**Fig. 2** Flow chart of the modeling algorithm

A uniform random number $\xi_1 \in (0\ 1]$ is used to determine the internal or residence time.

$$\Delta t_{int} = -\frac{\ln\xi_1}{\omega} \tag{10}$$

In the conventional KMC algorithm, an event is carried out in each simulation iteration and the timescale is advanced by $\Delta t_{int}$. However, $\Delta t_{int}$ in some iterations

may become extremely large. Therefore a second timescale, $\Delta t_{ext}$, is introduced and is referred to as the external time scale. $\Delta t_{ext}$ is a simulation parameter. The magnitude of $\Delta t_{int}$ relative to $\Delta t_{ext}$ determines whether a transformation event will occur or not in the present simulation step.

$$
\begin{cases}
\text{if } \Delta t_{int} > \Delta t_{ext} : t = t + \Delta t_{ext} \text{ and no event occurs} \\
\\
\text{if } \Delta t_{int} \leq \Delta t_{ext} : t = t + \Delta t_{int} \text{ and an event occurs}
\end{cases}
\tag{11}
$$

When $\Delta t_{int} > \Delta t_{ext}$, no event occurs and the timescale is advanced by $\Delta t_{ext}$. This allows one to update the boundary conditions in a more continuous manner and also capture the critical transformation stresses or temperatures. When $\Delta t_{int} \leq \Delta t_{ext}$, an event occurs and the one to be carried out is selected using the activation rate array as is normally done in KMC. The activation rates for all possible events are calculated using Eqs. (7-8) and are strung together to a total length $\omega$, with each event occupying a distance on the string equal to their individual activation rate. A second uniform random number $\xi_2 \in (0\ 1]$ is generated, and the vector from the origin to a distance $\xi_2 \omega$ along the string points at the event to be executed. After the transformation event, time is increased by $\Delta t_{int}$ and boundary conditions are updated.

## 5    FEM Modeling of Unit Transformation and Stress Distribution

### 5.1    Boundary Conditions

To simulate stress-induced or thermally-induced transformation, the applied stress or the temperature should change with time. They could be updated in each simulation step using the KMC time $t$ determined in Eq. (11). For example, an applied uniaxial stress $\sigma_a(t)$ is

$$
\sigma_a(t) = \sigma_{max} - |\dot{\sigma}_0 t - \sigma_{max}| \ (0 \leq t \leq 2\sigma_{max}/\dot{\sigma}_0)
\tag{12}
$$

where $\dot{\sigma}_0$ is a stress ramping rate and $\sigma_{max}$ is the maximum stress on the triangular loading curve common in experiments. Both loading and unloading segments are considered in Eq. (12). Under thermal cycling, the temperature $T$ first decreases at a ramping rate $\dot{T}_0$ from an initial high temperature $T_i$ to a low temperature $T_e$ and then returns to $T_i$. The temperature is varied in each simulation step according to

$$
T(t) = T_e + |\dot{T}_0 t + (T_e - T_i)| \ \left(0 \leq t \leq 2\left(T_i - T_e\right)/\dot{T}_0\right).
\tag{13}
$$

During thermal cycling, $\sigma_a$ can be either kept at zero, in which case thermally induced transformations are studied, or held at a constant value, which can lead to stress-assisted two-way shape memory effect. The transformation latent heat could

be assumed immediately exchanged with the environment when appropriate, and in this case the sample has a uniform temperature $T$ everywhere.

It is desirable that the boundary conditions be kept simple. For example, when a sample is subject to a uniaxial load, the stress at its top surface could be updated in each simulation step according to Eq. (12) while keeping its bottom surface constrained in the loading direction and its side surfaces free without any constraint.

## 5.2     Application of a Unit Transformation

Displacive martensitic transformation in a domain is implemented as a mandatory, stress-free distortion. The system subsequently relaxes by partitioning and balancing stresses with the surrounding region. This is analogous to the determination of the back stress in the classical Eshelby problem demonstrated in Fig. 3a. In the absence of an external stress, an elliptical region is subject to a mandatory positive shear and then the system is allowed to equilibrate and relax. Positive shear stress develops ahead of the vertices of the ellipse as these regions deform in order to accommodate the shear in the elliptical region, while negative shear stress (in blue) develops inside the sheared/transformed region due to the resistance to shear in the surrounding region. Figure 3b shows the shear stress distribution after a hexagonal region transforms when the far-field resolved shear stress is about 270 MPa. Despite the positive far-field shear stress, the shear stress in the transformed martensite region is negative (in blue), leading to a decrease in the activation energy barrier for reverse transformation, $E_{MM \to A}$ (Eq. (5)). On the other hand, the shear stress immediately outside the transformed region is positive and is very high, resulting in a reduced energy barrier for forward transformation,
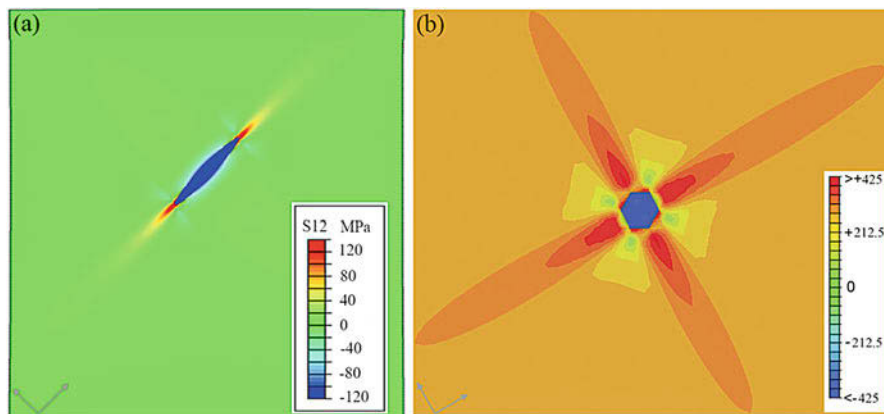


**Fig. 3** (**a**) Shear stress (in MPa) distribution after the system relaxes from a stress-free mandatory shear in an elliptical region along the major axis of the ellipse in the absence of an external stress. (**b**) Shear stress (in MPa) distribution after the system relaxes from transformation in a hexagonal region at a finite external stress

$E_{A \to MM}$ (Eq. (5)). In the subsequent simulation step, there will be competition between reverse transformation of this martensite region and forward martensitic transformation in the surrounding regions.

After each transformation event, the redistributed stresses are then used to calculate new activation rates and the time increment in KMC. Each simulation step ends with an updated boundary condition, and in the case of $\Delta t_{int} \leq \Delta t_{ext}$, a new phase configuration as well.

## 6     Example

### 6.1     Martensitic Transformation in Cu-Al-Ni SMAs

In this section, we will demonstrate the model using Cu-Al-Ni SMA as a model material. In this material, austenite has a $D0_3$ crystal structure (see Fig. 4a) and is often denoted as $\beta_1$. The deformation gradient matrices for lattice dilation and distortion associated with martensitic transformation from $\beta_1$ to $\beta_1{}'$ are provided in Eqs. (14) and (15), respectively, both of which are expressed for the martensite coordinate system shown in blue in Fig. 4a.

$$F^1 = \begin{bmatrix} 1.062 & 0 & 0 \\ 0 & 1.0232 & 0 \\ 0 & 0 & 0.918 \end{bmatrix} \tag{14}$$

$$F^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0.1154 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{15}$$

From Eq. (14) the volumetric strain (i.e., $F_{11}^1 \cdot F_{22}^1 \cdot F_{33}^1 - 1 = -0.25\%$) associated with the transformation is determined to be negligibly small. On the other hand, Eq. (15) leads to a large simple shear with a magnitude of 11.54%. Therefore, for simplification, transformation is implemented as a simple shear $\gamma_0 = 0.1154$ on $\{\bar{1}10\}$ planes and along <110> directions in austenite. In Eq. (7), the transformation unit volume $\Omega_0$ is taken as 6.5 $nm^3$ and the attempt frequency $\nu_0$ is taken as $2 \times 10^{10}$ $s^{-1}$. In other materials where martensitic transformation also involves large dilation and volumetric strain, the full strain tensor should be taken into consideration.

A challenge with modeling martensitic transformation lies in the lack of information on mechanical properties of the martensite phase. The example demonstrated in this section assumes that elastic properties for austenite and martensite are the same and that the model material is a two-dimensional single-crystalline solid under a plane strain assumption. The austenite phase is an elastic solid with cubic (nonisotropic) symmetry and therefore has three independent elastic constants: $C_{11} = 143.1$ GPa, $C_{12} = 124.3$ GPa, and $C_{44} = 94.1$ GPa (Sedlák et al. 2005), which lead
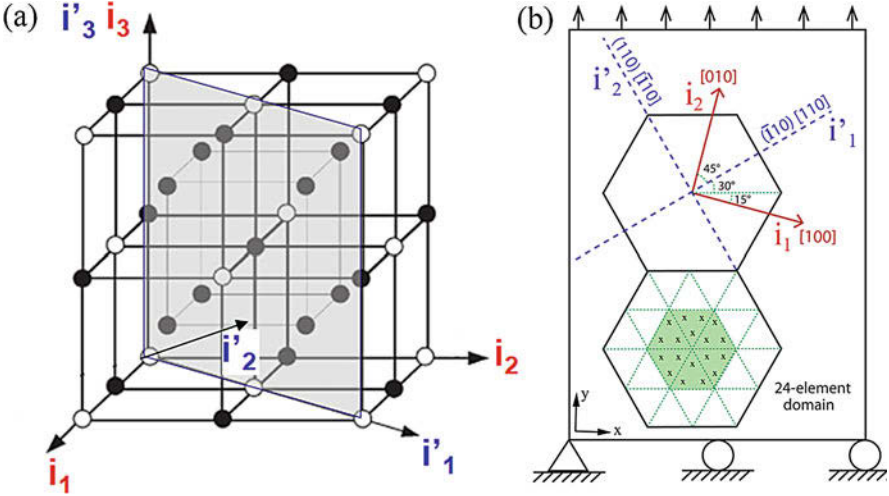
**Fig. 4** (**a**) The D0$_3$ crystal structure for the Cu-Al-Ni alloy. ○ denote Al atoms, • at the edge center and body center (1/2, 1/2, 1/2) are mostly Cu atoms and some are Al atoms, while • in the center in each sub-unit are mostly Cu atoms and some are Ni atoms. The deformation gradient matrices in Eqs. (14-15) are written for the i'$_1$, i'$_2$, i'$_3$ martensite coordinate system. (**b**) Model setup in the example. The two-dimensional sample is fixed at the bottom left corner, constrained vertically at the bottom, and subjected to a uniaxial strain $\varepsilon_a(t)$ at the top surface. Its side surfaces are free, while its top and bottom surfaces are constrained as a plane. Each transformation domain is a hexagon whose lines of symmetry (dashed lines) correspond to the transformation system {1$\bar{1}$0} < 110>. Each domain is meshed with 24 elements, among which the six in the center (the crosses are their integration points) are used to calculate the average resolved shear stress for this domain

to Young's modulus $E = 27.5$ GPa, the Poisson's ratio $\nu = 0.4648$, and the shear modulus $G = 94.1$ GPa. Plasticity is not considered as martensitic transformation usually occurs at stresses much lower than the yield stress. However, local yielding is possible, and this effect could be an interesting direction for future extension of the model.

## 6.2    Computational Considerations

Partition of the simulation cell into nonoverlapping domains is desirable as each location in the material is associated with a phase identity at any given time. Each domain may be further meshed with one or more elements in the FEM implementation. The progression of the transformation is simulated as successive transformation in different domains, and the material volume involved in each transformation event could be allowed to vary from event to event depending on the spatial stress distribution. In this section, the demonstrations are from a

simple hexagonal tessellation of the material; each hexagonal domain is a phase transformation unit, and only a single domain is involved in each transformation event. The model setup is illustrated in Fig. 4b. Each domain is further meshed with 24 triangular plane strain, 6-node quadratic elements, and stresses at the three integration points inside each element are solved. Due to the noncircular domain geometry, stress concentration develops near domain edges and corners (see Fig. 3b). To exclude such (unphysical) stress concentration spots, only the six elements in the center of each domain are used in the calculation of the average resolved shear stress in a domain. With this approach, mechanical interactions across phase boundaries can be captured by using a sufficiently small domain size; in principle, the domain size should be much smaller than the sample and characteristic scale of the evolving microstructure in order to accurately capture the pattern in phase and stress evolution.

To avoid severe local distortion near or at the corners of the specimen, a top layer and a bottom layer are prohibited from transforming. Domains in these layers are set up as austenite initially, and their transformation rates ($\omega_{A \to MM}$) are assigned as zero regardless of their local stress state. They never transform as a result, providing both elastic accommodation and constraint for the transforming material in between and promoting side surfaces as nucleation sites.

The stored elastic energy in the sample, $E_{el}$, in Eq. (3) is incorporated indirectly. For example, consider a potential forward transformation in a domain that would result in high stresses in its surroundings and therefore a very high $E_{el}$ and a high $E_{ise}$. Instead of raising the energy barrier $E_{A \to MM}$ in Eq. (5) to hinder the transformation, the model may allow the domain to transform. But it will likely revert to austenite immediately because the driving force for reverse transformation will be high and $E_{MM \to A}$ will be low due to the resulting high local stress $\tau$ ($\tau$ is incorporated in the evaluation of transformation energy barriers as shown in Eq. (5)). This approach is computationally much more efficient than determining $E_{el}$ for each domain, which requires carrying out every possible transformation and calculating the resulting change in the strain energy of the sample. Therefore, $E_{ise}$ for each domain is reduced to $\gamma_i A_i$, the change in austenite/martensite interface energy. An alternative approach is to evaluate $E_{el}$ analytically using some kind of volumetric average of Eshelby predictions for multiple scenarios based on neighboring phase identities. This approach is in principal similar as the approach above but with a modified value of $\gamma_i$.

It is typical for the KMC model to get trapped in the "superbasin" composed of many frequent events with much lower energy barriers for a large number of iterations before the system experiences a rare process and escapes out of the "superbasin" (Chatterjee and Voter 2010). In the present model, frequent and rapid phase flipping may occur during transformation in either direction when there is some sort of metastable phase configuration. It is reasonable to impose a limit $N_{AL}$ on the number of attempts allowed at each stress level. A transformation event occurs only if $\Delta t_{int} \leq \Delta t_{ext}$ and $N_{attempt} \leq N_{AL}$; otherwise $t$ will be increased by $\Delta t_{ext}$ to increase the driving force for transformation in one direction in subsequent steps.

## 6.3    Superelasticity (Mechanically-Induced Transformation)

In this section, modeling results for transformation under a constant strain rate are presented for the simplified case of a single martensite variant. This boundary condition is used frequently in experimental studies that explore rate effects in SMAs (Otsuka et al. 1976; Ueland et al. 2012; Van Humbeeck and Delaey 1981; Yin et al. 2013) and also allows one to capture possible stress relaxation behavior. Under a constant strain rate, the transformation should occur gradually over a period of time because transformation strain increases with martensite fraction.

Snapshots of the phase evolution and the corresponding shear stress distribution are shown in Fig. 5 for both the forward transformation (top two rows in Fig. 5) and the reverse transformation (bottom two rows in Fig. 5). Martensitic transformation starts in the center of the sample. Under a constant strain rate, formation of long plates that span across the sample leads to a higher overall transformation strain and more stress relaxation compared to shorter plates at corner regions that are constrained by the buffer layers at the top and the bottom of the sample, which may play a similar role as unfavorably oriented grains that do not transform within the range of the applied load. The corner regions partially transformed at higher strains (e.g., $\varepsilon = 0.04$-$0.052$) and retained some austenite until very high applied strains. In other regions, new martensite plates formed at relatively low strains and grew and impinged at higher strains. Formation of a martensite plate usually leads to extensive stress relaxation. During reverse transformation, the corner regions reverted to austenite first, and formation of austenite plates results in stress spikes.

The corresponding superelastic stress-strain curve obtained is presented in Fig. 6a. Frequent stress fluctuations are seen during both forward and reverse transformations in Fig. 6a showing sudden stress drops (rises) during forward (reverse) transformation, each followed by a period of elastic loading (unloading) with a slope parallel to that of the initial elastic loading curve. Numerous small serrations are seen in Fig. 6b, where martensite fraction $f_M$ is plotted as a function of strain $\varepsilon$. Overall, martensite fraction $f_M$ still seems linearly related to strain $\varepsilon$ (e.g., linear fitting of the forward transformation data in Fig. 6b yields $f_M = 20.61(\varepsilon - 0.0268)$), but a linear relationship is no longer accurate because $f_M$ increases in a discontinuous manner. $f_M$ actually increases or decreases by a small percentage at a constant $\varepsilon$ and then pauses over a range of $\varepsilon$ before increasing or decreasing at another constant $\varepsilon$. The insets in Fig. 6b show how the overall external stress $\sigma$ changes with $f_M$. During forward transformation, the increase in $f_M$ always results in a decrease in $\sigma$, which subsequently increases at a constant $f_M$. Therefore, with increasing applied strain, strain increment is first achieved by elastic deformation of the sample, leading to an increase in $\sigma$ and a reduction in the free energy of martensitic state, $G_{MM}$. Subsequently, martensitic transformation occurs in some regions of the sample and the resulting transformation strain fulfills the strain increment requirement, leading to relaxation of the elastic stress developed prior to transformation. Stress relaxation has indeed been observed in tensile testing of small SMA microwires under a constant slow strain rate (Ueland et al. 2012). During reverse transformation, the decrease in $f_M$ always leads to
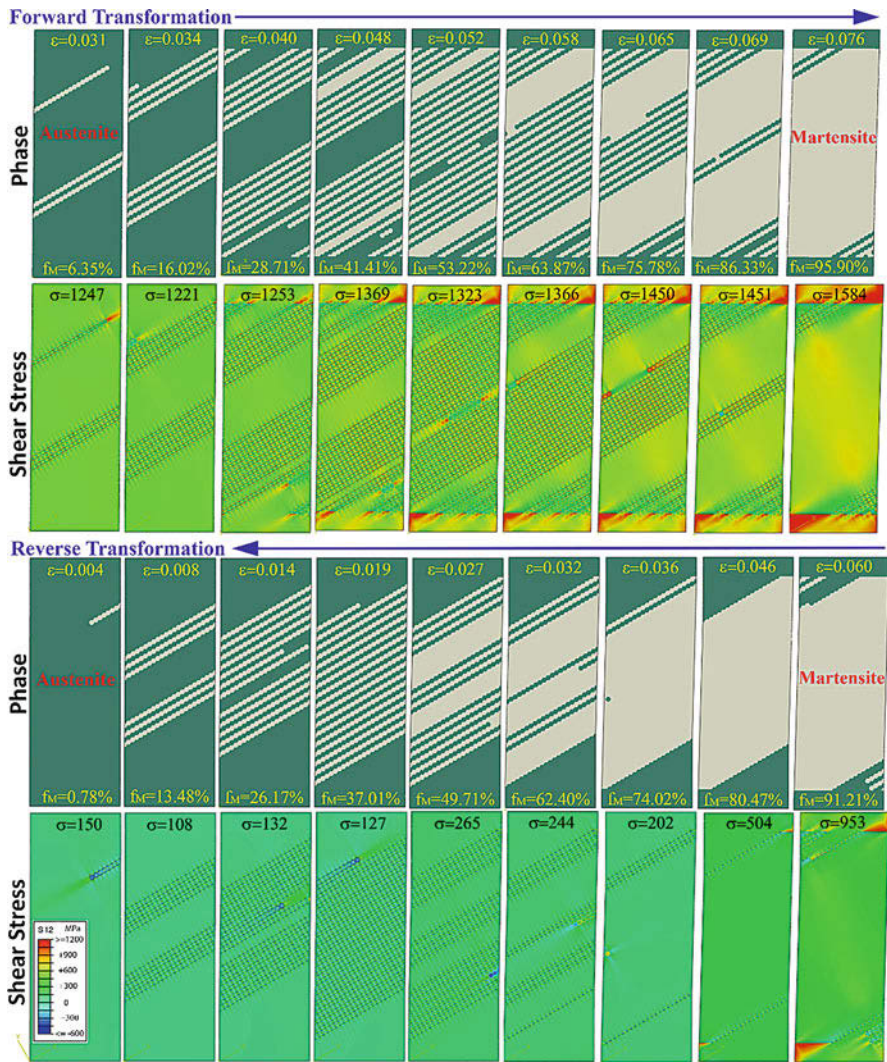
**Fig. 5** Phase configurations (austenite in green and martensite in beige) and the corresponding resolved shear stress distributions during forward transformations (top two rows of figures) and reverse transformations (bottom two rows of figures) in a mechanical cycle under a constant strain rate. The strain $\varepsilon$ and stress $\sigma$ (in MPa) values in the figure titles are uniaxial strain and stress values for the sample. $f_M$ is the volume fraction of martensite

an increase in $\sigma$, which subsequently decreases at a constant $f_M$. Strain reduction is first realized by elastic unloading, which in turn raises the driving force for reverse transformation. Eventually reverse transformation takes place, meeting the strain reduction requirement by the recovered transformation strain. The stress rises rapidly and then the next iteration begins.

**Fig. 6** (**a**) Superelastic stress-strain curve showing complete strain recovery after unloading. Transformations are accompanied by stress fluctuations. (**c**) Martensite fraction $f_M$ as a function of strain. The insets plot $f_M$ as a function of stress; forward and reverse transformations lead to load drops and increases, respectively
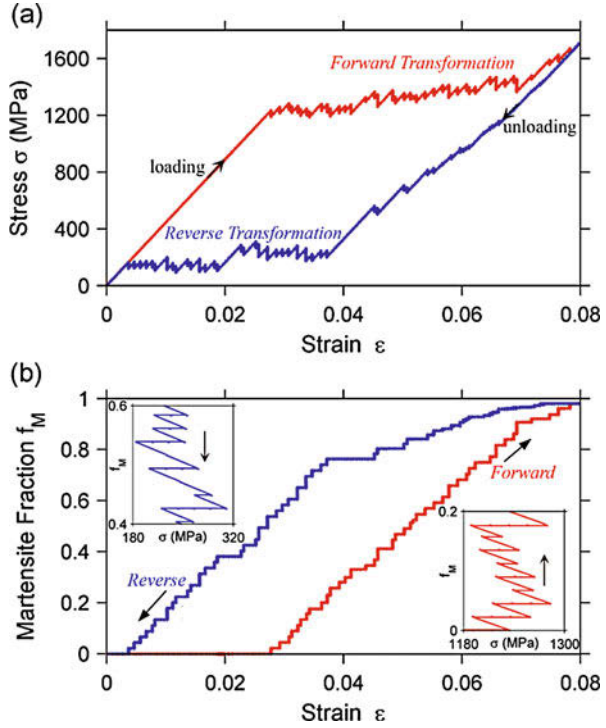
## 6.4     Thermally-Induced Transformation

The modeling framework presented in Sects. 2, 3, 4, and 5 can also be applied to simulate thermally-induced transformation, either in the absence of an external load or at a constant load. The example in Fig. 7 below shows simulation results for forward and reverse transformation in a thermal cycle when there is no applied load. The material properties are the same as specified above in Sect. 6.1. For simplicity, only two martensite variants with opposite transformation shear strains are permitted. The equilibrium temperature $T_0$ is set to 500 K. The temperature $T$ is set to 800 K initially and then is gradually decreased to 200 K before it is increased back to 800 K.

The fractions of two martensite variants, $f_M^1$ and $f_M^2$, are plotted as a function of the temperature $T$ in Fig. 7a. During cooling, $f_M^1$ and $f_M^2$ remain zero until $T$ decreases to about 300 K when $f_M^1$ and $f_M^2$ increase simultaneously to slightly above 0.4. The phase configurations during cooling are shown in Fig. 7b–e. It can be seen in these figures that plates of the two martensite variants nucleate simultaneously and grow in alternating patterns. As the two variants have opposite transformation strains, the alternating pattern in Fig. 7b–e enables self-accommodation among the variants and minimizes the macroscopic strain in the sample. Both variants

**Fig. 7** (**a**) The fractions of martensite variant 1 and variant 2, $f_M^1$ and $f_M^2$, as a function of temperature under no external stress. $T_0$ is the equilibrium temperature which is set as 500 K. (**b**–**e**) Phase configurations during cooling, showing transformation from austenite (in green) to martensite (the two variants are plotted in red and beige)
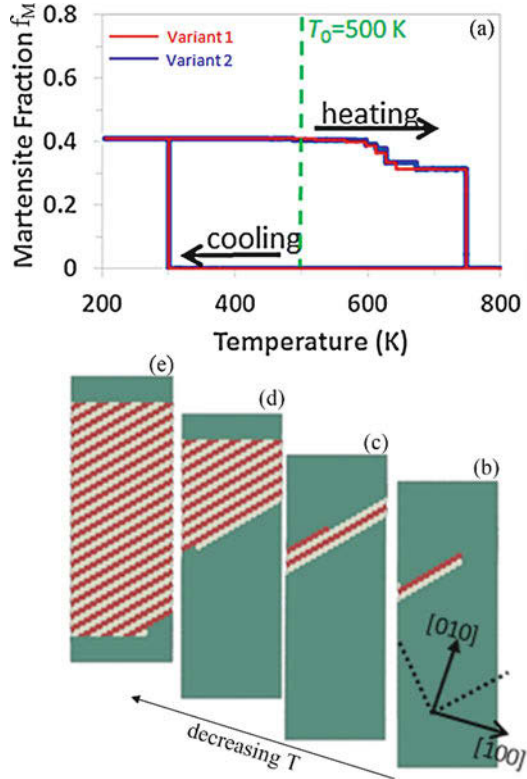
transform back to austenite upon heating and the reverse transformation is complete at about 750 K, yielding a temperature hysteresis. Transformation temperatures and thermal hysteresis are normally measured experimentally by Differential Scanning Calorimetry (DSC), which however cannot predict evolution in phase configuration. The modeling framework presented here offers a complementary method for studying thermally-induced martensitic transformation.

## 7 Conclusions

This chapter presents a mesoscale kinetic Monte Carlo (KMC) model framework for modeling the dynamics of reversible martensitic transformation in shape memory alloys and ceramics. The model may also be adapted to study martensitic transformation in structural materials in which case incorporation of plastic deformation mechanisms is essential. The modeling framework integrates KMC with thermodynamic description of transformation and Finite Element (FEM) prediction of mechanical states. The thermodynamic energetics governing martensite transfor-

mations induced by temperature changes or an applied stress are examined, and the energy landscapes for transformation states are developed. Thermodynamics is related to kinetics through a rate equation that is commonly used to describe thermally-activated phenomena. KMC provides FEM a new phase distribution as well as a new timescale (and thus new boundary conditions). The thermomechanical distribution determined from FEM in turn leads to changes in the energy landscape for the transformation states.

Computational implementation of the model is realized and presented. Martensitic transformation occurs by a mandatory displacement, which is mainly a shear, and is modeled as a unit process that takes place in one domain at a time. After each unit transformation event, the boundary conditions are updated according to the new timescale and the new stress states in the sample are calculated by FEM. The new phase configuration and stress distributions are used to determine the activation rate for each domain based on a transition state approximation for the energy landscape of the transformation states. The KMC algorithm is then used to select the next domain to transform and determine the time increment. This iterative simulation method therefore performs virtual in-situ testing under mechanical, thermal, and combined stimuli and predicts simultaneous evolution in phase and thermal-mechanical states in the sample using physical time and length scales comparable to those involved in experiments. It in particular captures inherent transformation stochasticity during both nucleation and growth stages.

The capability of the model is demonstrated by using it to simulate mechanically-induced reversible transformation (superelasticity) and thermally-induced transformation (in which case multiple variants are introduced) in plane strain two-dimensional specimens. Although not demonstrated, the model can also easily simulate mechanically-induced transformation followed by reverse transformation during heating (one-way shape memory effect) as well as transformations induced by thermal cycling at a constant stress (two-way shape memory effect). The model predicts phase and stress evolution during the entire incubation, nucleation, and growth process of the new phase. It also predicts (rather than enforces) both forward and reverse transformations with a stress or temperature hysteresis and with complete recovery. The predicted critical transformation stresses or temperatures are related to fundamental thermodynamic properties of materials. Stress relaxation and spikes are observed during transformations under a constant applied strain rate.

This modeling approach could significantly augment mesoscale computational capabilities that connect microstructures to macroscopic thermomechanical properties. The model is capable of capturing dynamic mechanical interactions between microstructure and stochastic transformation quantitatively and addressing issues of transformation kinematics in the presence of defects. An important direction for future research is to elucidate various size effects and defect effects (e.g., grain boundary effects) on martensitic transformation utilizing the KMC modeling method described in this chapter.

# References

Bonnot E, Vives E, Mañosa L, Planes A, Romero R (2008) Acoustic emission and energy dissipation during front propagation in a stress-driven martensitic transition. Phys Rev B 78:094104

Bulatov VV, Argon AS (1994a) a stochastic-model for continuum elastoplastic behavior.1. Numerical approach and strain localization. Model Simul Mater Sci Eng 2:167–184. https://doi.org/10.1088/0965-0393/2/2/001

Bulatov VV, Argon AS (1994b) a stochastic-model for continuum elastoplastic behavior. 2. A study of the glass-transition and structural relaxation. Model Simul Mater Sci Eng 2:185–202. https://doi.org/10.1088/0965-0393/2/2/002

Bulatov VV, Argon AS (1994c) a stochastic-model for continuum elastoplastic behavior. 3. Plasticity in ordered versus disordered solids. Model Simul Mater Sci Eng 2:203–222. https://doi.org/10.1088/0965-0393/2/2/003

Chatterjee A, Voter AF (2010) Accurate acceleration of kinetic Monte Carlo simulations through the modification of rate constants. J Chem Phys 132:194101. https://doi.org/10.1063/1.3409606

Chen Y, Schuh CA (2011) Size effects in shape memory alloy microwires. Acta Mater 59:537–553. https://doi.org/10.1016/j.actamat.2010.09.057

Chen Y, Schuh CA (2015) A coupled kinetic Monte Carlo–finite element mesoscale model for thermoelastic martensitic phase transformations in shape memory alloys. Acta Mater 83:431–447. https://doi.org/10.1016/j.actamat.2014.10.011

Gall K, Sehitoglu H (1999) The role of texture in tension–compression asymmetry in polycrystalline NiTi. Int J Plast 15:69–92. https://doi.org/10.1016/S0749-6419(98)00060-6

Gall K, Lim TJ, McDowell DL, Sehitoglu H, Chumlyakov YI (2000) The role of intergranular constraint on the stress-induced martensitic transformation in textured polycrystalline NiTi. Int J Plast 16:1189–1214. https://doi.org/10.1016/s0749-6419(00)00007-3

Guda Vishnu K, Strachan A (2012) Size effects in NiTi from density functional theory calculations. Phys Rev B 85:014114

Homer ER, Schuh CA (2009) Mesoscale modeling of amorphous metals by shear transformation zone dynamics. Acta Mater 57:2823–2833. https://doi.org/10.1016/j.actamat.2009.02.035

Homer ER, Rodney D, Schuh CA (2010) Kinetic Monte Carlo study of activated states and correlated shear-transformation-zone activity during the deformation of an amorphous metal. Phys Rev B 81:064204 https://doi.org/10.1103/PhysRevB.81.064204

Jaworski A, Ankem S (2005) The effect of α phase on the deformation mechanisms of β titanium alloys. J Mater Eng Perform 14:755. https://doi.org/10.1361/105994905x75565

Jin YM, Artemev A, Khachaturyan AG (2001) Three-dimensional phase field model of low-symmetry martensitic transformation in polycrystal: simulation of $\zeta'2$ martensite in AuCd alloys. Acta Mater 49:2309–2320. https://doi.org/10.1016/S1359-6454(01)00108-2

Juan JS, No ML, Schuh CA (2009) Nanoscale shape-memory alloys for ultrahigh mechanical damping. Nat Nanotechnol 4:415–419

Karaca HE, Karaman I, Basaran B, Lagoudas DC, Chumlyakov YI, Maier HJ (2007) On the stress-assisted magnetic-field-induced phase transformation in Ni2MnGa ferromagnetic shape memory alloys. Acta Mater 55:4253–4269. https://doi.org/10.1016/j.actamat.2007.03.025

Kastner O, Eggeler G, Weiss W, Ackland GJ (2011) Molecular dynamics simulation study of microstructure evolution during cyclic martensitic transformations. J Mech Phys Solids 59:1888–1908. https://doi.org/10.1016/j.jmps.2011.05.009

Kriven W (1995) Displacive transformations and their applications in structural ceramics. J Phys IV 5:C8-101–C8-110

Lagoudas DC (2008) Shape memory alloys: modeling and engineering applications. Springer, Boston

Lagoudas D, Hartl D, Chemisky Y, Machado L, Popov P (2012) Constitutive model for the numerical analysis of phase transformation in polycrystalline shape memory alloys. Int J Plast 32-33:155–183. https://doi.org/10.1016/j.ijplas.2011.10.009

Levitas VI, Levin VA, Zingerman KM, Freiman EI (2009) Displacive phase transitions at large strains: phase-field theory and simulations. Phys Rev Lett 103:025702

Levitas VI, Roy AM, Preston DL (2013) Multiple twinning and variant-variant transformations in martensite: phase-field approach. Phys Rev B 88:054113

Mamivand M, Asle Zaeem M, El Kadiri H, Chen L-Q (2013) Phase field modeling of the tetragonal-to-monoclinic phase transformation in zirconia. Acta Mater 61:5223–5235. https://doi.org/10.1016/j.actamat.2013.05.015

Mamivand M, Asle Zaeem M, El Kadiri H (2014) Phase field modeling of stress-induced tetragonal-to-monoclinic transformation in zirconia and its effect on transformation toughening. Acta Mater 64:208–219. https://doi.org/10.1016/j.actamat.2013.10.031

Manchiraju S, Anderson PM (2010) Coupling between martensitic phase transformations and plasticity: a microstructure-based finite element model. Int J Plast 26:1508–1526. https://doi.org/10.1016/j.ijplas.2010.01.009

Manchiraju S, Gaydosh D, Benafan O, Noebe R, Vaidyanathan R, Anderson PM (2011) Thermal cycling and isothermal deformation response of polycrystalline NiTi: simulations vs. experiment. Acta Mater 59:5238–5249. https://doi.org/10.1016/j.actamat.2011.04.063

Ortín J, Planes A (1988) Thermodynamic analysis of thermal measurements in thermoelastic martensitic transformations. Acta Metall 36:1873–1889

Otsuka K, Wayman CM (1998) Shape memory materials. Cambridge University Press, Cambridge

Otsuka K, Wayman CM, Nakai K, Sakamoto H, Shimizu K (1976) Superelasticity effects and stress-induced martensitic transformations in Cu-Al-Ni alloys. Acta Metall 24:207–226

Otsuka K, Sakamoto H, Shimizu K (1979) Successive stress-induced martensitic transformations and associated transformation pseudoelasticity in Cu-Al-Ni alloys. Acta Metall 27:585–601

Ozdemir N, Karaman I, Mara NA, Chumlyakov YI, Karaca HE (2012) Size effects in the superelastic response of $Ni_{54}Fe_{19}Ga_{27}$ shape memory alloy pillars with a two stage martensitic transformation. Acta Mater 60:5670–5685. https://doi.org/10.1016/j.actamat.2012.06.035

Patoor E, El Amrani M, Eberhardt A, Berveiller M (1995) Determination of the origin for the dissymmetry observed between tensile and compression tests on shape memory alloys. J Phys IV 5:C2-495–C2-500

Sedlák P, Seiner H, Landa M, Novák V, Sittner P, Mañosa L (2005) Elastic constants of bcc austenite and 2H orthorhombic martensite in CuAlNi shape memory alloy. Acta Mater 53:3643–3661

Tadaki T, Otsuka K, Shimizu K (1988) Shape memory alloys. Annu Rev Mater Sci 18:25–45

Tao K, Choo H, Li H, Clausen B, Jin J-E, Lee Y-K (2007) Transformation-induced plasticity in an ultrafine-grained steel: an in situ neutron diffraction study. Appl Phys Lett 90:101911

Tatar C, Kazanc S (2012) Investigation of the effect of pressure on thermodynamic properties and thermoelastic phase transformation of CuAlNi alloys: a molecular dynamics study. Curr Appl Phys 12:98–104. https://doi.org/10.1016/j.cap.2011.04.050

Ueland SM, Chen Y, Schuh CA (2012) Oligocrystalline shape memory alloys. Adv Funct Mater 22:2094–2099. https://doi.org/10.1002/adfm.201103019

Van Humbeeck J, Delaey L (1981) The influence of strain-rate, amplitude and temperature on the hysteresis of a pseudoelastic Cu-Zn-Al single crystal. J Phys Colloq 42:C5-1007–C5-1011

Wechsler MS, Lieberman DS, Read TA (1953) On the theory of the formation of martensite. Trans AIME J Met 197:1503–1515

Wollants P, Roos JR, Delaey L (1993) Thermally- and stress-induced thermoelastic martensitic transformations in the reference frame of equilibrium thermodynamics. Prog Mater Sci 37:227–288

Wood AJM, Clyne TW (2006) Measurement and modelling of the nanoindentation response of shape memory alloys. Acta Mater 54:5607–5615. https://doi.org/10.1016/j.actamat.2006.08.013

Yin H, Yan Y, Huo Y, Sun Q (2013) Rate dependent damping of single crystal CuAlNi shape memory alloy. Mater Lett 109:287–290. https://doi.org/10.1016/j.matlet.2013.07.062

Zhong Y, Gall K, Zhu T (2012) Atomistic characterization of pseudoelasticity and shape memory in NiTi nanopillars. Acta Mater 60:6301–6311. https://doi.org/10.1016/j.actamat.2012.08.004

# Object Kinetic Monte Carlo (OKMC): A Coarse-Grained Approach to Radiation Damage

# 58

Christophe Domain and Charlotte S. Becquart

## Contents

C. Domain (✉)
EDF R&D, Département MMC, Les Renardières, Moret sur Loing, France
EM2VM, Joint laboratory Study and Modeling of the Microstructure for Ageing of Materials, Lille/Moret sur Loing, France
e-mail: christophe.domain@edf.fr

C. S. Becquart
UMET, Unité Matériaux et Transformations, University Lille, CNRS, INRA, ENSCL, UMR 8207, Lille, France
EM2VM, Joint laboratory Study and Modeling of the Microstructure for Ageing of Materials, Lille/Moret sur Loing, France
e-mail: charlotte.becquart@univ-lille.fr

## Abstract

Object kinetic Monte Carlo (OKMC) approaches allow one to explicitly coarse grain atomic processes to simulate the evolution with time of the system/microstructure. This class of methods is quite versatile and can be used to model different processes where the motion of atoms in a given microstructure and their interaction with sinks and traps of the microstructure lead to changes such as phase transition or modification. In this chapter, the method is presented as applied to radiation damage modeling. Along with a description of the technique, we discuss the different pathways possible to couple OKMC with smaller- as well as larger-scale methods and conclude with a brief enumeration of what we believe are the issues for future development.

## Abbreviations

| | |
|---|---|
| AKMC | Atomic kinetic Monte Carlo |
| BKL | Bortz, Kalos, and Lebowitz |
| DFT | Density functional theory |
| EKMC | Event kinetic Monte Carlo |
| F/M | Ferritic/martensitic |
| FIA | Foreign interstitial atoms |
| FP | Frenkel pair |
| GPU | Graphics processing unit |
| KMC | Kinetic Monte Carlo |
| MD | Molecular dynamics |
| MFRT | Mean field rate theory |
| NEB | Nudged elastic band |
| ODS | Oxide dispersion strengthened |
| OKMC | Object kinetic Monte Carlo |
| PBC | Periodic boundary conditions |
| PD | Point defects |
| RPV | Reactor pressure vessel |
| RTA | Residence time algorithm |
| SFT | Stacking fault tetrahedras |
| SIA | Self-interstitial atom |
| ST | Self-trapping |
| TM | Trap mutation |

# 1  Introduction

Kinetic Monte Carlo (KMC) is a family of methods based on Markov/stochastic processes intended to simulate the time evolution of some processes occurring in nature, typically processes that occur with a given known rate, used as inputs to the

algorithm (Voter 2007). This method is typically applied at two scales: the atomistic scale, where the processes that can take place are the ones atoms undergo during the evolution of a microstructure, and the mesoscopic scales, where the microstructure is coarse grained into objects that can evolve on a lattice or not. It is now clearly established that high-energy particles such as the ones produced in a nuclear reaction interact with matter and lead to nuclear reactions, displacement of atoms, as well as ionization and excitation of atoms (Was 2017). In the case of metallic alloys, the displacement of atoms is the main aging mechanism in irradiated components (pressure vessel, internals, or cladding for fission reactors; divertor or blanket for fusion reactors) as the displacements of atoms lead to the formation of point defects (PD), the evolution of which with time will induce changes in the microstructure. The interaction between radiation-induced point defects and solute atoms leads to solute atoms and point defect fluxes which drive the microstructure evolution with time and induce changes in the mechanical properties. Microstructure modeling techniques dedicated to radiation damage studies thus focus on the evolution of these point defects. In fact, these techniques, as applied currently, aim at modeling the microstructure evolution under irradiation by looking at the point defect and point defect cluster evolution and distributions. At the mesoscopic scale, the species under consideration are defects whose atomic-level features are disregarded, except insofar as they influence the mechanisms, whereby the defects migrate or react between themselves and with sinks (Was 2017). The two techniques the most used are mean field rate theory (MFRT) (Ghoniem) and object kinetic Monte Carlo (OKMC), which is the subject of this chapter. OKMC methods sit at the center of the microstructure multiscale modeling picture as they have links with lower-scale methods (e.g., atomistic ones) and upper-scale ones (phase field, MFRT). Indeed, OKMC coarse grain lower-scale atomic simulations and can be coarse grained into MFRT or plasticity/mechanics modeling.

The chapter is structured as follows: we will start by briefly describing what is meant by microstructure evolution under irradiation and the modeling techniques used for that purpose. We will then present the KMC method as it is applied to objects and describe the ingredients of the models. In the fourth section, we will discuss different approaches to model alloys as well as ways to couple the OKMC to lower- and higher-scale techniques. Finally, in the fifth section, we conclude by providing what we believe are future directions and perspectives of the technique. Some of the existing OKMC codes are briefly presented in the Appendix A.

## 2   Radiation Damage Evolution

### 2.1   Physical Processes Involved in Radiation Damage

Radiation damage is caused by high-energy particles, neutrons, or ions, penetrating materials and interacting with their nuclei or their electrons. As mentioned briefly in the introduction, the interaction can lead to nuclear reactions (transmutations) and thus the formation of new elements, ionization and excitation of atoms, as well as

atomic displacements. The first two outcomes (atom transmutation and excitation) need to be modeled at the scale of nuclei or electrons and will thus not be discussed here. The third one (atomic displacements) results in the creation of a large amount of point defects, vacancies, and self-interstitial atoms (SIAs), isolated or in clusters within displacement cascades. This is referred to as the primary damage. These point defects are created in overconcentration and flow into the materials and interact with each other as well as solute atoms and defects of the microstructure such as dislocations, grain boundaries, or surfaces. The primary damage evolution into voids and dislocation loops is the so-called matrix damage. Besides the matrix damage, the point defect oversaturation has other consequences. First, it can lead to a radiation-accelerated diffusion making the system return faster to its equilibrium state. Second, because of the tendency of point defects to eliminate on sinks (e.g., dislocations, grain boundaries, surfaces), the concentration of PD near the sinks is lower than the average concentration which results in concentration gradients inducing a flow of PD.

As the microstructural changes induced by the atomic displacements result from the evolution of the point defects that have been created, methods aiming at modeling the microstructure evolution focus on point defect motions rather than atoms. Different methods are available which are briefly described in the next section.

## 2.2    Simulation Techniques of Irradiated Microstructures

The formation of the primary damage, referred to as the "source term" in mesoscale methods such as KMC or MFRT, occurs too quickly ($\sim$10–20 ps), and the displacement cascade sizes are generally too small, to be observed experimentally, but molecular dynamics (MD) simulations have proven to be a powerful tool to model displacement cascades and investigate the creation and morphology of the defects created, their mutual interaction, and, to some extent, i.e., at very short times, also their evolution (Becquart et al. 2018). However, because of the short timespan covered by MD simulations (10–20 ps for volumes large enough to contain a few displacement cascades, i.e., of the order of tens of nanometers), other computational tools must be used to extend the study up to the formation of experimentally resolvable damage features. One natural way of speeding up the calculations while keeping the information at the atomistic scale is the atomic kinetic Monte Carlo approach (AKMC) (Becquart and Soisson 2018; Becquart and Domain 2010). The AKMC method is based on the residence time algorithm (RTA), also known as the BKL method (Bortz et al. 1975) (for Bortz, Kalos, and Lebowitz) developed by Young and Elcock to model the evolution of vacancies in binary ordered alloys (Young and Elcock 1966). This algorithm is used to evolve a set of atoms and point defects in a simulation box by point defect jumps $i$, occurring with thermally activated frequencies. AKMC simulations (see recent reviews in Becquart and Domain 2010; Soisson et al. 2010; Becquart and Wirth 2012) have shown to be especially useful to model systems of a few million atoms on the appropriate

timescale, because they deal with the jumps of atoms or point defects, not the lattice vibrations. AKMC is thus a powerful technique to study the microstructural and microchemical evolution of alloys controlled by diffusion processes in all kinds of conditions (e.g., thermal aging, irradiation) (Becquart and Soisson 2018). This approach naturally takes into account the correlations between successive jumps, inherent to the process of diffusion (Allnatt and Lidiard 2003). A consequence is that a very large number of jumps have to be performed to model the motion of a point defect cluster as shown, for instance, in Athènes and Bulatov (2014) in the FeCu system. To overcome this problem, mesoscale approaches of the KMC method, such as the OKMC, have been developed. OKMC is defined as a coarse-grained method because atoms are not explicitly treated in contrast with atomistic KMC. It is also based on the residence time algorithm applied this time to objects: the microstructure is described in terms of objects which are the intrinsic defects (vacancies and self-interstitials) or impurities and their clusters which are located at known (and traced) positions in a simulation volume. The events are all the possible actions that these objects can perform and the reactions that they may undergo, such as migration (in most cases – isotropic diffusion – a random walk with small jumps corresponding to nearest neighbor distance in the crystal), dissociation (emission of a smaller defect from a bigger one), clustering of like defects or of defects and impurities, and annihilation between opposite defects (self-interstitials and vacancies). The probability for a migration event is given by the appropriate jump frequency.

Very similar coarse-grained KMC models are the event KMC (EKMC) (Lanore 1974) which do not treat individual migration jumps explicitly and where the evolution of the system is driven by reactions (events) between objects.

Finally note that mean field rate theory (MFRT) is another mesoscopic approach largely used to simulate the evolution of microstructure under irradiation. MFRT models do not deal with the explicit position of defects but with the averaged concentration of defects in a small volume at a given position in the material. They take into account the presence of gradients of concentration and hence, in some sense, deal with the spatial dependence of the evolving defect distribution.

Figure 1 presents the different methods currently used to model material microstructure evolution under irradiation. Each of the phenomena involved in the microstructure or plasticity evolution has a characteristic time and length scale that range from nanometers and picoseconds for the displacement cascades up to micrometers, for the grain sizes and hours to years for the experiment duration or the lifetime of components. Consequently one needs to choose the appropriate modeling technique that is the most efficient or at least capable of simulating one of these phenomena.

## 2.3    Microstructure Modeling by Kinetic Monte Carlo Techniques

As stated in the introduction, KMC techniques aim at modeling the behavior of point defect and point defect clusters which, for the most common cases, migrate in the
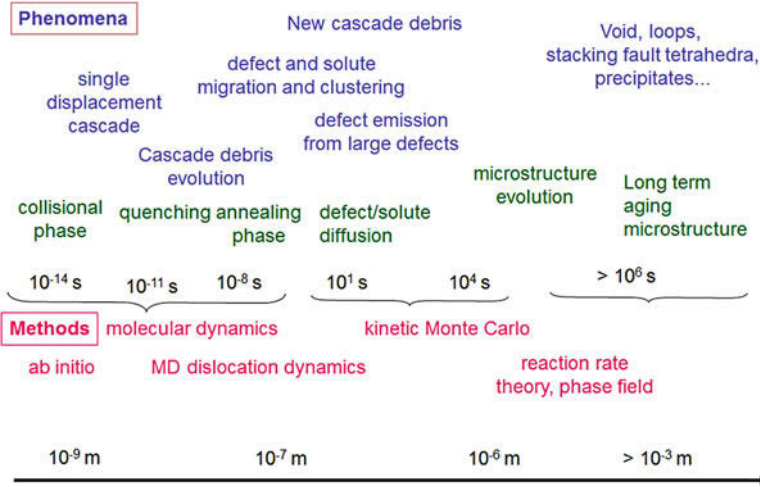
**Fig. 1** Different methods, physical phenomena, timescales, and length scales required to model the evolution of microstructures under radiation damage conditions

material via the jumps of point defects. We now present the method as it is applied in AKMC, i.e., when all the atoms are explicitly taken into account, for the sake of simplicity. For a thermally activated process such as point defect diffusion, the jump frequency is given by:

$$\Gamma = \nu e^{-\frac{E_a}{k_B T}} \tag{1}$$

where $\nu$ denotes the attempt frequency and T the temperature and $E_a$ corresponds to the activation energy of the jump or the migration energy obtained as the difference between the energies of the system at the saddle point and in its initial configuration. Such a description is derived from the theory of thermally activated processes and is justified when the thermal fluctuations are smaller than the activation energies: $k_B \ll E_a$.

According to harmonic transition state theory (Vineyard 1957), the attempt frequency can be expressed as:

$$\nu = \frac{\prod_{j=1}^{3N-3} \nu_j}{\prod_{j=1}^{3N-4} \nu_j^*} \tag{2}$$

In this expression $\nu_j$ and $\nu_j^*$ are the normal frequencies for vibrations at the local minimum and saddle states, respectively, and $N$ is the number of atoms. Typical values for the migration of isolated point defects in metals are around $10^{12}$–$10^{13}$ s$^{-1}$ ($\sim$ Debye frequency).

Knowing the frequencies of all possible defect jumps in the system, the algorithm proceeds by choosing one of them according to its probability and evaluate the time necessary for the jump to occur using the residence time algorithm (RTA)

that will be described in the next section. Originally applied to the migration of vacancies in binary alloys, the AKMC method can be applied to any point defects as well as foreign interstitial atoms (FIA) such as C, N, He, and H (that do not need point defects to diffuse) and substitutional solute atoms. The microstructure evolves via elementary mechanisms, i.e., point defect jumps or solute jumps in the case of interstitial solutes. Vacancies and self-interstitials can jump from one lattice site to another lattice site (typically first nearest neighbor sites); solute interstitials lie on an interstitial sublattice and jump on this sublattice. Typically one atom is assumed to move, and the rates of possible transitions are determined from the local environment around the moving atom. During the course of the simulation, the probabilities of all the possible jumps are calculated, and one migrating event is chosen, at each time step, by extracting a random number, according to its probability.

## 3   OKMC Method and Principles

The OKMC method proceeds as the AKMC, except that instead of making atoms and point defects evolve individually, clusters of point defects evolve as a whole as one single object. Figure 2 compares the two approaches, whereas the main differences between OKMC and EKMC are represented in Figs. 3 and 4.

Diffusion can be treated as a random walk on a lattice or off lattice in OKMC and as a direct reaction in EKMC (Fig. 3). Consequently, in OKMC, reactions are treated when objects are below a specific interaction distance, whereas all possible reactions are considered and computed in EKMC (Fig. 4). Schematically, in OKMC, all events are elementary events as exposed below, whereas in EKMC they are coarse grained in reaction events. One consequence is that 1D or anisotropic diffusion is easy to implement in OKMC.

### 3.1   General Principles

The objects treated in the OKMC model have defined center-of-mass positions in space, which can correspond to lattice sites or not, and (when they are supposed to) they migrate. They can be point defects and point defect clusters, interacting with solute atoms to form mixed complexes, traps that can be introduced to simulate
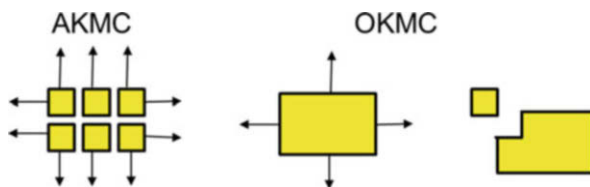


**Fig. 2** Simple comparison between AKMC and OKMC. The arrows are the possible moves of each square which represents, for instance, a single vacancy for AKMC and the same 6 vacancy clusters for OKMC that can also undergo an emission event

**Fig. 3** Simple comparison between OKMC and EKMC for the diffusion of the small object (the large object is considered as immobile). In OKMC, the reaction can occur after a random walk of the small object, i.e., as a series of elementary reactions. The only event is the reaction of the small object with the big one in EKMC. (See Appendix A for more details on the EKMC)
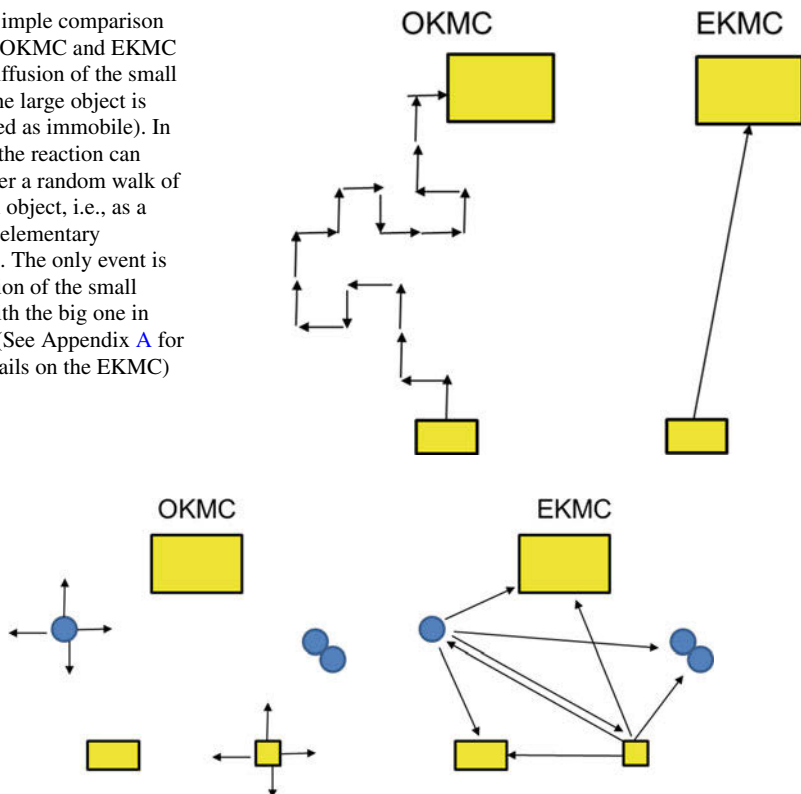


**Fig. 4** Simple comparison between OKMC and EKMC considering only the small objects as mobile (monovacancy in yellow and mono-SIA in blue). For OKMC the arrows represent the different jumps and for EKMC the different reactions. (See Appendix A for more details on the EKMC)

the effect of interstitial impurities and other objects of the microstructure such as dislocations and grain boundaries (Fig. 5). Depending on the level of detail required, the objects can have different shapes. For instance, they can be 3D spheres for nanovoids, 2D platelets for SIA loops, and stacking fault tetrahedra (SFT) for the defects found in face-centered cubic metals (Domain et al. 2004; Nandipati et al. 2015).

Each object and microstructure feature has a size, and an interaction volume (Fig. 6), equivalent to a distance between objects that takes into account the shape of the objects. In most cases the elastic interactions are taken into account implicitly, and effective interaction volumes/sizes are considered that are slightly larger than the object's real size. Depending on the code, two objects interact when their reaction volume overlap or when they touch each other. The interaction volume is directly related to the object sink strength (Jansson et al. 2013) which characterizes the interaction of migrating species with mobile or immobile features
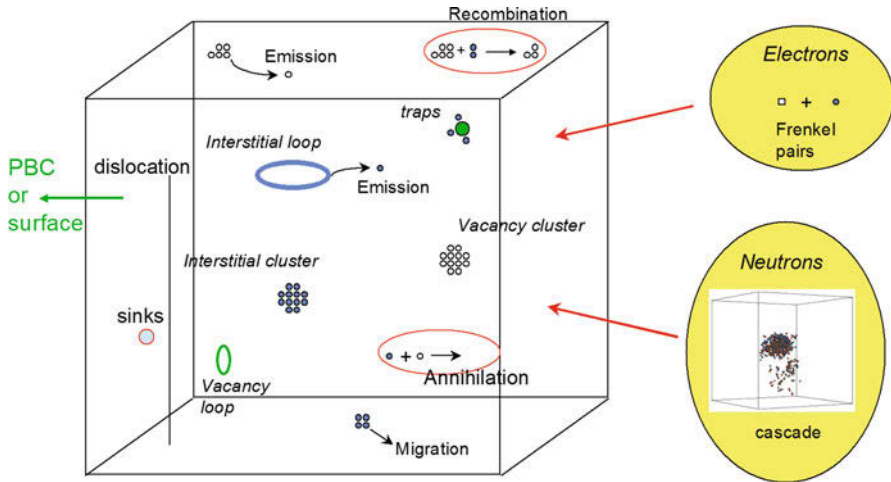
**Fig. 5** OKMC representation of a typical microstructure under irradiation



| Interaction between two spherical objects | Interaction between a loop and a spherical object |
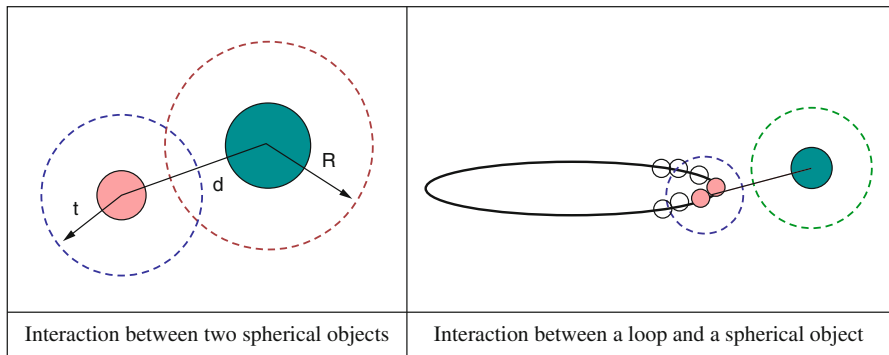
**Fig. 6** Object interaction for different shapes. Typical values for single point defects in transition metals are R $\sim$ 5 Å, and the core radius of a dislocation is about 5 Å

of the microstructure (Sect. 4.5.2). Note that elastic effects can be taken into account more specifically, as discussed in Sect. 5.

During the simulation, events can occur, whose probabilities of occurrence are expressed in terms of frequencies (in s$^{-1}$). From the set of all possible events $i$ of each object $n$ among the N objects in the system, each with a known rate of occurrence, $\Gamma_{n,i}$, a total rate, $R$, is computed:

$$R = \sum_{n,i} \Gamma_{n,i} \tag{3}$$

A Monte Carlo step is initiated by selecting an event from all the events by drawing a random number between 0 and $R$. The value of this number determines

which event is selected. If more than one object can undergo the selected event, one is chosen at random. Once the event and the corresponding object are selected, the appropriate actions are performed to make that event occur, and the time is updated, by adding a time increment proportional to the inverse of $R$.

The associated time-step length $\delta t$ and average time-step length $\Delta t$ is given by:

$$\delta t = \frac{-\ln r}{R} \quad \Delta t = \frac{1}{R} \tag{4}$$

where $r$ is a random number in [0,1].

## 3.2    Mechanisms and Events Occurring During Radiation Damage

The most common events that need to be considered in an OKMC model of irradiation are:

- **Migration**, i.e., jump to a neighboring site. The defects execute random diffusion jumps (in one, two, or three dimensions depending on the nature of the defect) with a probability proportional to their diffusivity, which has to be known for all mobile species.
- **Recombination** with objects of opposite type (i.e., SIA with vacancies) or aggregation with other objects (e.g., point defect clusters and complexes including solute atoms and FIAs).
- **Dissociation/emission** (when the object size is larger than 2) of one element (e.g., a single vacancy or single SIA) from the object.
- **Trapping** of objects at traps to model in a simple way microstructure features or heterogeneities such as impurities or specific solutes (e.g., carbon in steels) or precipitates (e.g., carbides in steels).
- **Annihilation** at free surfaces, grain boundaries, dislocations, or other sinks.
- **Trap mutation**, a common event observed in metals filled with noble gas (most often helium atoms (alpha particles) in fusion and fission conditions for structural materials and Xe or Ar in uranium oxide fuel pellets). Helium atoms are strongly repelled by metallic atoms and as consequence they have a propensity to form clusters. When too many He atoms are aggregated together, the system needs to relieve the strain created by the interstitial elements by the ejection of one or more matrix atoms leading to the formation of one or more Frenkel pairs (FP), i.e., vacancies and SIAs. As He binds very strongly to vacancies, the He cluster will be trapped by the vacancy it created and will not be capable of moving. This is thus a self-trapping (ST) event as the binding energy of the helium cluster with the vacancy is usually quite large. If one or more vacancies are already associated with the He cluster, the same mechanism can take place which is usually referred to as trap mutation (TM) or loop punching, if more than one SIA is created (when many SIAs are emitted at the same time, they are usually bound together also and form a dislocation loop). These FP creations make He clusters and He vacancy
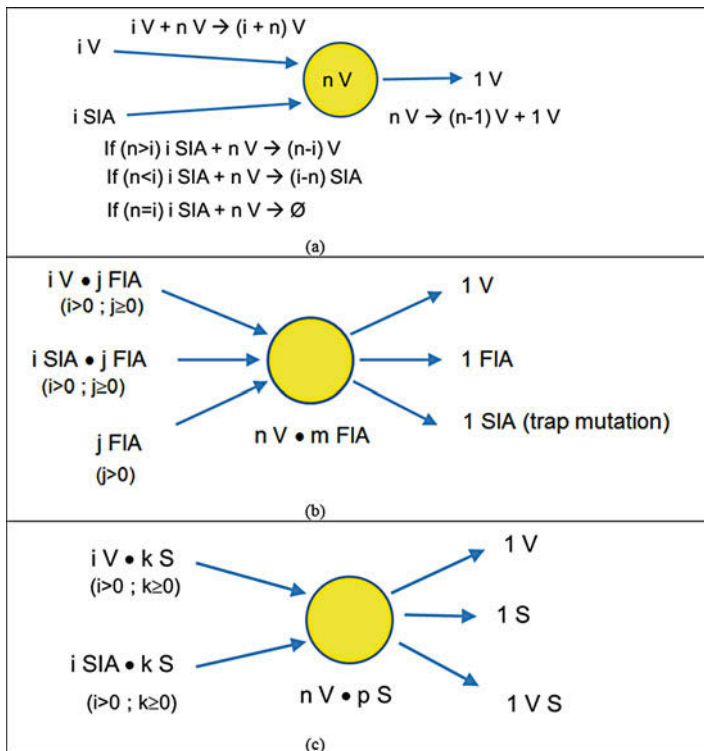
**Fig. 7** (**a**) Possible reactions at a vacancy cluster containing n vacancies. (**b** and **c**) Events (reaction and emission) at a complex object consisting of n vacancies, m foreign interstitial atoms (FIA), or p solute atoms (S): nV•mFIA or nV•pS

(nHe.mv) clusters be nonsaturable traps for themselves unlike what is observed for hydrogen atoms, for instance. Self-trapping and trap mutation consist thus in the simultaneous creation of a vacancy and a SIA; the latter remains bound to the He clusters (De Backer et al. 2015).

Figure 7 summarizes the main different recombination and dissociation events an object can undergo for (a) a vacancy cluster (nV), (b) a complex vacancy-FIA cluster (nV.mFIA), and (c) a complex-solute cluster (nV.pS). This schematic graph shows the increase in complexity when a FIA or a solute is associated to point defects and point defect clusters and the necessity to define a much larger number of events (and parameterize them).

The simulation of irradiation can be treated as the occurrence of *external events*, which then depend upon the kind of irradiation that is simulated: to simulate electron irradiation Frenkel pairs are introduced at random with the proper rate; for a neutron irradiation, displacement cascade debris are introduced also at random with the proper rate. For an ion implantation simulation, one has to take into account the

implantation profile, and the cascade debris are thus introduced according to the profile. All external events are characterized by rates $P_m$ (for event m), corresponding to the production rate ($P_m$ = (number of external events/cm$^3$/s) × (simulation box volume)). Since each of these external events occurs at a known rate, their rates are added to the computation of the total rate.

Jump, dissociation/emission, detrapping and trap mutation are considered as thermally activated processes, characterized by an activation energy $E_{a,n,i}$ and an attempt frequency $\nu_{n,i}$. Their rates, for object n and event i are obtained as:

$$\Gamma_{n,i} = \nu_{n,i} \exp\left\{-\frac{E_{a,n,i}}{kT}\right\} \tag{5}$$

For the object jump, the activation energy is its migration energy, $E_m$. For the emission from either a trap or an object, the most typical method is to take the activation energy $E_a$ as the sum of the binding energy, $E_b$, of the escaping entity (e.g., that of a single vacancy or SIA from a cluster) and its migration energy $E_m$; $E_a = E_m + E_b$. Note also that in a real diffusion event, on a real lattice, the particle emitted has to perform several jumps to escape the source object. This can be accounted for by placing the escaping object at a distance such that, at the following step, both recombination with the initial object and diffusion away are possible; thus $\Gamma_{n,i}$ is only an approximation of the escaping probability. If one wants to compare directly with rate equation models, the emitted particle is placed at a random position in the simulation box (Stoller et al. 2008).

Trapping and annihilation of defects with opposite defects or at sinks, as well as aggregation into larger clusters, take place spontaneously whenever the involved objects come to a mutual distance smaller than a reaction distance, which is equal to the sum of the capture radii associated to each of the two objects. The most typical sinks are free surfaces, grain boundaries, and dislocation lines (segments). As a small part of matter can be simulated using OKMC, specific shapes can be used for the simulation box in order to reproduce experimental samples as illustrated in Fig. 8.

Dissociation and emission events have an activation energy often taken to be equal to the sum of the migration energy $E_m$ plus the binding energy between the emitted defect and a cluster of the size that remains after emission, $E_b$. The corresponding rate is:

$$\Gamma_d = \nu_0 \exp\left(-\frac{E_m + E_b}{k_B T}\right) \tag{6}$$

Regarding the reaction between defects, it is generally assumed that they are diffusion limited. That is, the reaction occurs as soon as the two objects are within a pre-defined capture radius, but no energy barrier is associated with the reaction. These events occur therefore only on the basis of geometrical considerations (overlap of reaction volumes) and do not participate in defining the progressing

| Bulk sample | Thin foil for transmission electron microscopes (TEM) | Atom probe specimen (or nanowire) |
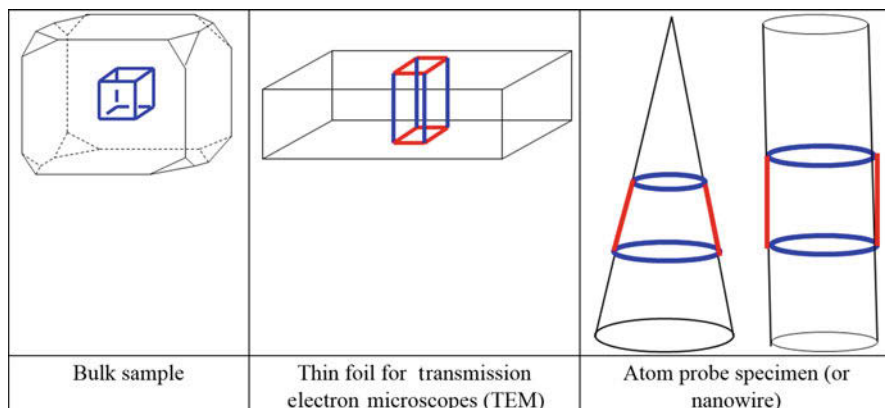
**Fig. 8** Specific sample shapes that can be simulated; periodic boundary conditions are applied on blue faces and free surfaces on red faces

of time. Events of this type also include the absorption of objects by sinks (e.g., impurities, dislocations, grain boundaries).

Grain boundaries, which act as sink for point defects, can be simply assimilated as free surfaces, as in Valles et al. (2017) where all the defects that reach the four lateral surfaces of the simulation box were removed out of the box or introduced using the average grain size method, inspired by Heinisch's work (Heinisch and Singh 2002) sometimes referred to as the finite PBC according to Soneda et al. (2003): when PBC are applied, whenever a mobile object has moved a distance larger than the average grain size, it is eliminated.

Dislocations can be introduced as segment(s) whose length is established according to the dislocation density to be simulated. In simple models, they act as linear sinks for point defects with a given recombination radius. Another approach could consist in considering dislocations as traps rather than sinks for specific objects, but this requires to then characterize the interaction mechanisms and the associated energies. Dislocation loops can be simulated by objects of toroidal shape, characterized by their major and minor radii and their orientation (Jansson et al. 2013).

The physical inputs of the model are thus the migration and attempt frequencies of all the objects that can form, their interaction volume, and the dissociation energies of all the possible objects that can be emitted.

# 4    Applications

## 4.1    Treatment of Impurities and Foreign Interstitial Atoms

One difficult issue in microstructure modeling is the comparison with experiments as real materials are far from being perfect and even the purest materials contain

impurities. One typical example is the long debate that existed regarding the migration energy of the vacancy in pure Fe, where two values were proposed all based on careful experiments. It was later shown that the highest value was the vacancy migration energy for Fe containing a slight amount of C acting as traps for the vacancy. Indeed, in a very pure Fe, Vehanen et al. (1982) found a vacancy migration energy of 0.55 eV, whereas Schaefer et al., in Fe containing a slight amount of impurities measured a value of 1.28 eV (Schaefer et al. 1977). One way to account for these impurities is to use effective parameters. For instance, one can use 1.3 eV as the migration energy of the vacancy in "real Fe" as real Fe will contain C as shown in an OKMC simulation of a vacancy diffusing in a pure Fe matrix (Becquart et al. 2010) containing immobile traps representative of C atoms. However, such an approach is quite limited as the concentration of the impurities is not taken into account. A more realistic method consists thus in introducing traps with appropriate binding energies and capture radius (Jansson and Malerba 2013, 2014). Note that this approach is limited to "slow" FIAs such as C and N in Fe or W at low enough temperatures. At temperatures where C and N can easily move or for fast-diffusing interstitials such as He and H in Fe and W, one needs to consider them as another object type that can move and form clusters, associated or not with point defects. In the case of rigid lattice approaches, the interstitial species will move on specific sublattices. The trap concentration can vary according to the aging or the dose, as done, for example, in Castin et al. (2018) to simulate transmutation and formation of Re in W alloys.

## 4.2 Taking into Account Solute Effects: "Gray Alloy" Approach of the Microstructure

The effect of solute atoms or alloying elements can be modeled using the so-called "gray alloy" approach. The "gray alloy" approach consists in a homogenization of the alloy composition. The alloy properties are described as in the case of pure metals by diffusion coefficients and binding energies characteristic of the alloy considered. More precisely, the migration energies, attempt frequencies, and binding energies of point defects and point defect clusters that locally depend on the environment (i.e., the number of solute atoms) of the moving species are considered to be the same everywhere (the local environment is neglected), and an effective value is considered. In this approach all the solutes can be treated in an effective way, or one can treat one solute (usually the one with the largest impact on the microstructure evolution) explicitly and all the other ones as part of the "gray alloy." A typical example is carbon in dilute Fe alloys. Because carbon interaction with vacancies is very large, carbon is usually treated explicitly, whereas Cr or Mn and Ni can be modeled as part of the "gray alloy." In this approach, one can have effective properties depending on the alloy composition. They are usually applied to the whole simulation box; however, it is possible to use it locally to model concentration gradients in a cellular approach of the method. The "gray alloy" approximation has been used, for instance, to model to predict the nanostructure

evolution under neutron irradiation in both reactor pressure vessel (RPV) and ferritic/martensitic (F/M) steels modeled, respectively, in terms of Fe-C-MnNi and Fe-C-Cr alloys (Chiapetto et al. 2015a, b). The effect of the substitutional solutes of interest was translated into modified parameters for the mobility of defect clusters. More precisely, the interstitial loop mobility was defined by an effective mobility which depends on the solute concentration. Using this approach, the origin of low-temperature radiation hardening (and subsequent embrittlement) was investigated, and the results support the hypothesis that solute clusters segregate on immobile interstitial loops, which act as heterogeneous nucleation sites for the formation of the NiSiPCr- and MnNi-enriched cluster populations (Chiapetto et al. 2015a). It was also shown that the introduction of Cr in F/M steels enhances recombination and suppresses the formation of vacancy voids. These results explain the link between the non-monotonic SIA diffusivity with %Cr and the experimentally observed radiation-induced swelling (Chiapetto et al. 2015b).

## 4.3   Taking into Account Solute Effects: Beyond the "Gray Alloy" Approach

In a "gray alloy" approach, the effect of solutes is introduced assuming that it changes the properties of the objects. This technique is rather simple and has been refined in a model that contains a more explicit description of solute atoms in dilute alloys (Castin et al. 2017). In this new approach, solute atoms are explicitly introduced in the matrix; however only single point defects are allowed to interact with them in first approximation. Point defect clusters still assume the gray matrix, i.e., they do not interact with the solute atoms, and their properties are effective properties depending upon the alloy composition. When either a single vacancy or a single SIA catches a solute atom (found within its radius of interaction), the latter is removed from the matrix and incorporated into the objects, thereby redefined. Another approach has been proposed (Domain et al. 2004) for low-content FeCu alloys: in the case of objects containing both point defects and solute atoms, the emission of either a single vacancy, a mixed vacancy-solute pair, or a single solute can be considered. In that later case, it is the remaining object that moves away from the emitted solute.

Solute concentration can be also accounted for using cellular KMC which seems to be particularly adapted for alloys. This method can be seen as the coarse graining of AKMC for solutes coupled to OKMC for defects, linked by transport coefficients. The system is decomposed in cells, defined by a local solute concentration, and solute exchange between neighboring cells is controlled by flux coupling with point defects. This method has been recently applied to Cu precipitation in Fe under thermal aging (Garnier and Nastar 2013) and spinodal decomposition in FeCr (Dopico et al. 2015; Rodríguez-Martínez et al. 2017). With this method direct simulation of the initiation of germination (for precipitation or spinodal decomposition) is not treated as the initial germ is smaller than the cell size. Possible solutions to this issue are to perform standard AKMC to initiate the

germination and switch to cellular KMC or initially reduce the size of the cells. The acceleration factor is found to be around two or three orders of magnitude for the cases investigated so far.

## 4.4  Downscale Coarsening: Coupling the AKMC with OKMC

The OKMC approach is sometimes used within a lower scale method as proposed in Castin et al. (2011) and later Pannier (2017) where the AKMC is coupled with the OKMC in the same code, thus producing a hybrid AKMC-OKMC code. In Castin et al. (2011), the precipitation of Cu in Fe was simulated using the following scheme: Cu clusters above 15 atoms are considered as objects, for which migration and dissociation events are defined and have specific, size-dependent, and thermally activated frequencies. Matching between the fully atomistic and the coarse-grained approach is achieved by using a neural network that provides all the necessary parameters for large Cu clusters, after a training on atomistically informed results. The same approach has been very recently developed to model the evolution of complex Fe alloys and more precisely the formation of solute-rich clusters representative of RPV steels, under irradiation (Pannier 2017). In this work, it was observed that about 80% of the jumps done in AKMC simulations of Fe alloys under irradiation concerned quasi-immobile clusters, the motion of which does not impact much the microstructure evolution and more precisely the formation of solute-rich clusters. Hence, losing details concerning the description of this type of object might not drastically change the stochastic trajectory of the simulation. As a consequence, such large clusters have been treated as objects. In the case of pure iron, when a group of first nearest neighbor point defects has a size (in number of point defects) higher than the user-defined object cutoff, all these point defects are considered as one object. The individual jump frequencies of the point defects associated to an object are null. Frequencies for migration and emission are associated with each object and obtained by a series of AKMC simulations. Migration is performed by moving the whole cluster by a first nearest neighbor distance and emission by putting the emitted point defect in a position near the object. When the size of an object changes, the object returns to the AKMC state. This AKMC relaxation is required because it allows the objects to return to an energetically more favorable configuration rather than being stuck in an atypical geometry. Figure 9 illustrates the hybrid AKMC-OKMC approach applied in Pannier (2017).

## 4.5  Upscale Coarsening: Using OKMC to Feed MFRT

### 4.5.1  Using the OKMC to Obtain the Source Term

As already mentioned in Sect. 2.2, MFRT is another possible approach widely used to model the evolution of microstructures under irradiation. A primary difference between the two methods is that spatial information about individual objects is maintained in OKMC, while only spatially averaged information is used in the
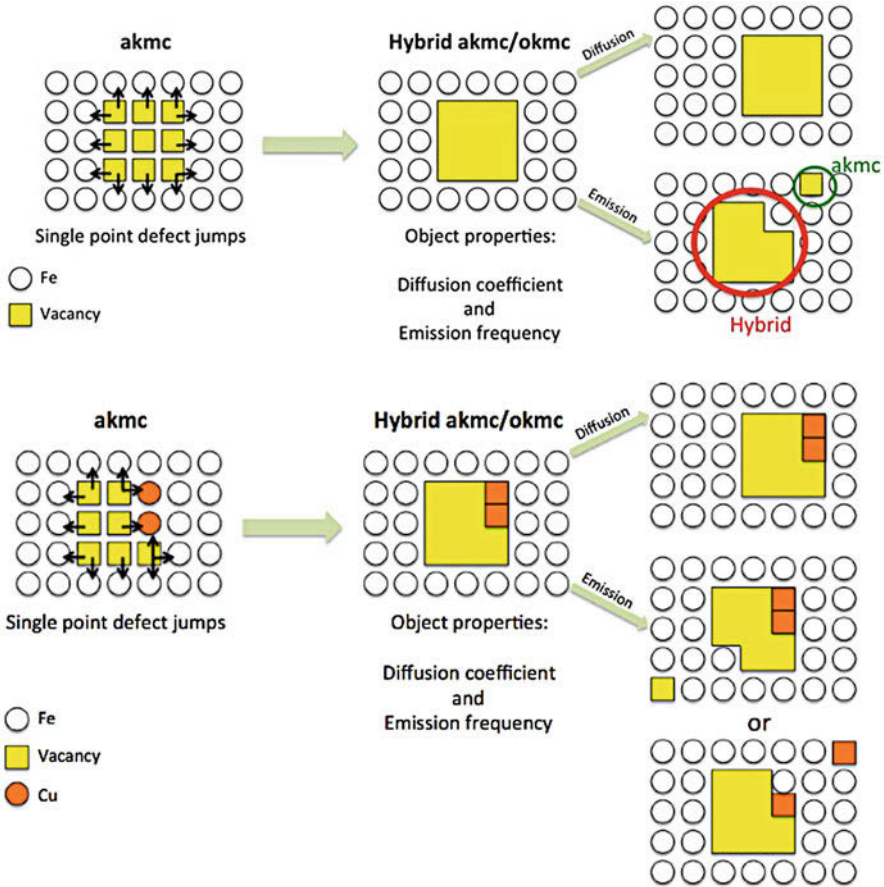
**Fig. 9** Illustration of the hybrid AKMC-OKMC method, for a 9-vacancy cluster and a 7-vacancy + 2 solute cluster (Pannier 2017). The 9-vacancy cluster (top) or the 7-vacancy and 2-solute cluster (bottom) are treated as single objects in the hybrid AKMC-OKMC. These two objects can then undergo two kinds of events. They can move as a block or emit a point defect or a solute. The emitted species are, in the next KMC step, treated in the AKMC approach. The remaining object, depending on its size, will be treated in the OKMC or in the AKMC manner. In the latter, the point defects and solutes will be treated individually

MFRT. Furthermore, OKMC algorithms deal with the explicit spatial dependence of the evolving defect distribution, whereas MFRT models do not deal with the explicit position of defects, but they deal with the averaged concentration of defects in a small volume at a given position in the material. MFRT models do however take into account the presence of gradients of concentration, and hence, in some sense, the spatial dependence of the evolving defect distribution is not fully lost. The source term however must be carefully estimated. Indeed, neutron or ion irradiations lead to the formation of cascade debris composed of isolated defects and defect clusters in

a confined region (of few 10 nm), and this spatial correlation has an important effect on the long-term evolution of the microstructure (Souidi et al. 2011). As MFRT cannot treat these spatial correlations, OKMC are often used to coarse grain the primary damage and obtain the source term (Heinisch 1983; Becquart et al. 2010; Adjanor et al. 2010; Jourdan and Crocombette 2012) for MFRT models.

### 4.5.2    Using OKMC to Evaluate Sink Strengths

The sink strength is a notion typically used in rate theory models to describe the interaction of migrating defects with the features characterizing the microstructure of the material (e.g., voids, dislocations, grain boundaries). The sink strength of each type of microstructural feature is proportional to the square of the inverse of the mean distance covered by the migrating species before it interacts with the feature, in general by being absorbed or trapped. It is a priori a function not only of the type, shape, orientation, size, and concentration of the sinks but also, and sometimes crucially, of the dimensionality of the motion of the migrating species (three-dimensional (3D) versus one-dimensional (1D) or mixed 1D/3D migrating defects). KMC techniques have the advantage of going beyond the mean field approximation, by explicitly and spontaneously taking into account spatial correlations between the elements of the physical system. They implicitly reproduce, among other phenomena, the effect of sinks or traps for migrating species, characterized by a given geometry and spatial distribution. They can thus provide directly the sink strengths of any microstructural features, provided, of course, that the feature is accurately introduced in the OKMC model (Heinisch et al. 2007; Malerba et al. 2007; Jansson et al. 2013; Hou et al. 2016; Ahlgren and Bukonte 2017; Carpentier et al. 2017).

## 5        Future Direction and Perspectives

Difficulties in applying the OKMC algorithm are not due to the method itself, which is fairly straightforward. They stem from the fact that all the possible events that each object can undergo according to a specific physical mechanism, their probability, and their properties must be pre-defined. The key point of this method is thus to determine the effective parameters from atomic models, for instance, as well as all the relevant mechanisms that can happen. It is important to stress that OKMC does not have the capability to predict structures or events that are not explicitly included in the model (unlike ab initio, molecular dynamics (MD) simulations or KMC methods with on-the-fly event catalog building capacity (Trochet et al. 2015; Henkelman and Jónsson 2001).

Another issue that needs to be stressed out is that the simulations are quite long and can last days to weeks. To parallelize efficiently the codes for complex microstructures is a challenge which has first been tackled to independent moving species (walkers) in an OKMC framework by Martínez et al. (2008). The same method has been applied more recently by Martin-Bragado et al. (2015) to OKMC, for a more realistic case of an isochronal annealing of Fe after electron irradiation.

Indeed, this kind of experiment can be simulated with such an algorithm as the initial distribution of point defects is homogeneous. There exists thus specific conditions where parallelizing is efficient as shown also in Jiménez and Ortiz (2016) where an OKMC code has been parallelized on GPU for a simple model considering only recombination and no "particle" clustering. However, for the time being, the efficiency of parallelizing OKMC simulations has not been proved for realistic situations, i.e., situations where objects with very different properties and in particular diffusion coefficients covering orders of magnitude have to be considered. To reach meaningful irradiation doses (>0.1 dpa) with realistic fluences, one still has to resort to homogenized rate theory methods based on the mean-field approximation (Marian et al. 2017). Furthermore, despite being capable of handling system sizes larger than the AKMC method, it is still difficult to model microstructures in which low-density objects or very large objects are formed in OKMC. At the moment, the order of magnitude of the concentration limit is around $10^{-15}$ cm$^{-3}$. Below this concentration, it is once again, more efficient to turn to MFRT techniques (Jourdan et al. 2014).

In the case of multicomponent alloys, the explicit treatment of all the solutes that can form a wide variety of complex objects composed of point defects and several solute types is also a challenge as the more objects, the more parameters in the OKMC model. Furthermore, the object shapes can also have some impact which adds another level of complexity in their treatment. These issues can benefit from AKMC simulations that can be used to obtain laws regarding the formation and interaction of these objects or other techniques such as the one proposed by Bonny et al. (2017). They developed a mean field model to obtain the energetics of point defect clusters in perfect random alloys and computed the dissociation energy of small vacancy clusters in FeCr and FeCr-2%W alloys, considered to be model alloys for Eurofer steels. They concluded that the dissociation energy is not expected to vary by more than 0.1 eV in the 0–10% Cr and 0–2% W composition range and thus that a mean field approach could be used in these situations.

The elastic effects can, in some cases, significantly bias the system evolution. An exemplary case is the interaction of dislocation loops with the surfaces of the thin foil samples used in TEM (Fikar et al. 2017a, b) or the interaction between point defects and semi-coherent interfaces (Vattré et al. 2016). OKMC is well-suited to incorporating long-range elastic interactions because the spatial location of defects can be tracked with high accuracy. One way proposed by Hudson et al. (2005) is to include them as a bias to the diffusive motion of mobile species. The work done by the elastic force over the course of the anticipated jump is calculated and used to bias the jump probability. Other, more explicit methods exist. One possibility is to take them into account when estimating the interaction between two objects as proposed by Mason et al. (2014) where it was shown that the experimental microstructure obtained by irradiating an ultrahigh purity with a very low fluence of self-ions could only be reproduced in OKMC if the elastic interactions between loops are taken into account. Another method, proposed by Vattré et al. includes the elastic interactions in the estimation of the activation barriers (Vattré et al. 2016). According to Carpentier et al. (2017), the impact of the elastic interactions on the

diffusion properties and their effect on the bias is mostly due to the anisotropy at the saddle point position. Elastic effects due to individual defects can also be evaluated by lower-scale methods such as MD as proposed in Chang et al. (2013, 2015) or continuous methods such as phase field (e.g., Rouchette et al. 2014). The corresponding sink strength and bias obtained with these different methods can then be introduced in OKMC simulations by considering, as mentioned in Sect. 3.1, effective sizes for the objects that take into account elasticity in an effective manner.

To conclude, OKMC are versatile and powerful methods to study complex systems with many different reactions and possibilities. The coupling between AKMC and OKMC is one way of taking into account the details of object structure when it matters (e.g., nuclei) as well as speeding up the evolution of large objects by coarse graining the individual defect jumps into a collective evolution. However, homogenization and further coarse graining is still necessary to reach the high doses and/or low-flux characteristics of situations encountered in nuclear reactors.

## A.1 Appendix: OKMC Codes Dedicated to Radiation Damage Studies

We now very briefly describe some of the codes developed and used for radiation damage studies in metals.

### A.1.1 JERK

The event kinetic Monte Carlo code JERK was one of the pioneering codes developed at CEA by Lanore (1974). In this approach the lattice is ignored, and the objects have continuous coordinates. The elementary events are binary encounters of two objects, emissions of a mobile object from a source such as a cluster or a surface. Probability laws are defined for each event. A typical simulation step consists of following the evolution of the object population over a macroscopic time interval. This is achieved by building the list of all possible events which may happen within the time interval, processing the events in the order of increasing delays and refreshing the event list after each event has been processed. When the event list is exhausted, the partners that did not react are displaced according to macroscopic diffusion laws (Barbu et al. 2005). This code has been used to model isochronal annealing in Fe (Mason et al. 2014) and FeC (Jourdan et al. 2011).

## A.1.2   BIGMAC

The kinetic Monte-Carlo code BIGMAC was developed at Lawrence Livermore National Laboratory by Johnson in collaboration with Caturla and Díaz de la Rubia (Johnson et al. 1998). This code has been used by different research groups to study many systems. BIGMAC tracks the locations of defects, impurities, and clusters as a function of time without relying on a lattice structure (it is however possible to modify the code to include a lattice if necessary). In addition, no hardwired reactions are introduced in the code, thereby allowing the user to directly define any type of system: this versatility allows the study of complex systems. The disadvantage of this approach appears for very high doses, where large defect clusters are formed, since it requires large memory to store the matrices for all defect reactions.

One key feature of BIGMAC is the recent implementation of the cellular approach (Caturla et al. 2016): the simulation box is divided into smaller boxes (cells) with different Cr content. The alloying element is thus treated in terms of average local concentrations, which are given (in atomic %) for each cell. Another peculiarity of the code consists in the fact that jump probabilities for the defects depend on the local solute concentration, as well as on the concentrations of neighboring cells. The cellular OKMC is a priori the ideal framework to introduce explicitly the dependence on local concentration of parameters that describe stability and mobility of point defects and their clusters, making its application to concentrated alloys such as high-Cr F/M alloys ideal.

## A.1.3   Lattice Kinetic Monte Carlo (LAKIMOCA)

LAKIMOCA code is a KMC as well as a Metropolis Monte Carlo package developed in standard C by Domain at EDF R&D in 1998. It has been extensively described in Domain et al. (2004). LAKIMOCA can take into account a large number of objects (defect clusters, solute atoms, vacancy-solute clusters, grain boundaries, dislocations, etc.) and exists in two versions: atomistic kinetic Monte Carlo (AKMC) or object kinetic Monte Carlo (OKMC), according to the nature of the object considered as mobile. The code allows easy introduction of different classes of immobile traps (mimicking trapping nano-features, as well as dislocations) and sinks (i.e., features like dislocations and grain boundaries where point defects and clusters can disappear), characterized by specific geometrical shapes (spheres, infinite cylinders, surfaces, etc.). The mobile objects (or more precisely their center of mass) are located on a lattice that can be bcc, fcc, and hcp and can jump from one lattice site to a first neighbor lattice site. The starting configuration can be created by the code itself with random distributions of initial defects and/or solute atoms or can be given by the user. LAKIMOCA can simulate damage production from electron, ion, and neutron irradiation. Neutron irradiation is simulated by a mixed flux of Frenkel pairs and defect cascade debris.

### A.1.4 Modular Monte Carlo (MMonCa)

MMonCa is an open-source kinetic Monte Carlo simulator developed by Bragado (Martin-Bragado et al. 2013). It is developed in C++ and distributed mainly for Linux (Ubuntu) sytems. MMonCa contains a Lattice KMC module, used mainly for simulation of epitaxy, and an off-lattice Object KMC module, used for simulation of damage irradiation in many materials. The presence of physical defects that diffuse in space implies the need to include diffusion as a transition rate and to define algorithms for space migration and particle interaction. The space is divided in small prismatic elements using a tensor mesh. Space is assumed to be homogeneous (in material, temperature, etc.) inside each small element. Each element obtains its material definition by calling a user-defined procedure that specifies the material structure in the simulation. In this way, very complex shapes containing different materials can be simulated. This approach is of particular interest in the case of concentrated alloys. Furthermore, the cellular OKMC algorithm initially developed in BIGMAC has been introduced in MMonCa, as well.

### A.1.5 KSOME

KSOME is a flexible and lattice based code patterned after the original FORTRAN code ALSOME (Heinisch 1983) developed by Heinisch. Objects of interest include vacancies, SIAs, interstitial impurities, and clusters of these defects. The objects are characterized by their type, size, shape, position, and orientation (Nandipati et al. 2015).

### A.1.6 McHERO

McHERO stands for Monte Carlo simulation of helium-bubble evolution and resolutions (Takahashi et al. 2010). It is based on kinetic Monte Carlo (KMC) techniques to simulate migration and coalescence of He bubbles and interactions with distinct three-dimensional geometric features of the material. The kinetics of bubble migration and coalescence include Brownian motion as well as bias-driven migration, such as in a stress or temperature gradient. The diffusion coefficients of helium bubbles are converted into probabilities of the occurrence of the events. The event is in this case the diffusion of a bubble within a time step $\delta t$. The probabilities are then summed up to find the total probability of the events. The total probability is used to normalize the probability of each event.

### A.1.7 MATEO

The hybrid atomistic/object KMC code MATEO is written in C++ and has been developed by Castin at SCK•CEN (Castin et al. 2017). Its OKMC approach is built upon the same philosophy and fundamental hypotheses (in terms of lattice-defect

and microstructure description, reproduction of neutron irradiation, etc.) as the LAKIMOCA code, but it has been refined in a model that contains a more explicit description of solute atoms (Castin et al. 2017). Solutes are explicitly introduced in the matrix, in an AKMC fashion, but only single point defects are allowed to interact with them. Point defect clusters still assume the gray matrix. When either a single vacancy or a single SIA catches a solute atom (found within its radius of interaction), the solute atom is removed from the matrix and incorporated into the object, which is then redefined into a new one. Finally, the solutes are reintroduced in the atomic matrix when the carrying point defect interacts with another object, i.e., either during a recombination event with an opposite kind of defect or merging with another defect of the same kind. Globally, solute transport is thus implemented in a way, which can be regarded as a compromise between providing atomistic-level description and keeping the evolution of the predicted microstructure from the gray alloy model.

# References

Adjanor G, Bugat S, Domain C, Barbu A (2010) Overview of the RPV-2 and INTERN-1 packages: from primary damage to microplasticity. J Nucl Mater 406:175–186. https://doi.org/10.1016/j.jnucmat.2009.09.006

Ahlgren T, Bukonte L (2017) Sink strength simulations using the Monte Carlo method: applied to spherical traps. J Nucl Mater 496:66–76. https://doi.org/10.1016/j.jnucmat.2017.09.006

Allnatt AR, Lidiard AB (2003) Atomic transport in solids. Cambridge University Press, Cambridge

Athènes M, Bulatov VV (2014) Path factorization approach to stochastic simulations. Phys Rev Lett 113. https://doi.org/10.1103/PhysRevLett.113.230601

Barbu A, Becquart CS, Bocquet JL et al (2005) Comparison between three complementary approaches to simulate 'large' fluence irradiation: application to electron irradiation of thin foils. Philos Mag 85:541–547. https://doi.org/10.1080/14786430412331334616

Becquart CS, Domain C (2010) Introducing chemistry in atomistic kinetic Monte Carlo simulations of Fe alloys under irradiation. Phys Status Solidi B 247:9–22. https://doi.org/10.1002/pssb.200945251

Becquart CS, Wirth BD (2012) Kinetic Monte Carlo simulations of irradiation effects. In: Konings RJM (ed) Comprehensive nuclear materials. Elsevier, Amsterdam, pp 393–410. https://doi.org/10.1016/B978-0-08-056033-5.00030-6

Becquart CS, Barbu A, Bocquet JL et al (2010) Modeling the long-term evolution of the primary damage in ferritic alloys using coarse-grained methods. J Nucl Mater 406:39–54. https://doi.org/10.1016/j.jnucmat.2010.05.019

Becquart CS, De Backer A, Domain C (2018) Atomistic modelling of radiation damage in metallic alloys. In: Hsueh CH et al. (eds), Handbook of Mechanics of Materials. Springer, Singapore pp 1–30. https://doi.org/10.1007/978-981-10-6855-3_21-1

Becquart CS, Soisson F (2018) Monte-Carlo simulations of precipitation under irradiation. In: Hsueh CH et al. (eds) Handbook of Mechanics of Materials. Springer, Singapore pp 1–29. https://link.springer.com/referenceworkentry/10.1007%2F978-981-10-6855-3_24-1

Bonny G, Castin N, Pascuet MI, Çelik Y (2017) Exact mean field concept to compute defect energetics in random alloys on rigid lattices. Phys B Condens Matter 517:25–29. https://doi.org/10.1016/j.physb.2017.04.032

Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. J Comput Phys 17:10–18. https://doi.org/10.1016/0021-9991(75)90060-1

Carpentier D, Jourdan T, Le Bouar Y, Marinica M-C (2017) Effect of saddle point anisotropy of point defects on their absorption by dislocations and cavities. Acta Mater 136:323–334. https://doi.org/10.1016/j.actamat.2017.07.013

Castin N, Pascuet MI, Malerba L (2011) Modeling the first stages of Cu precipitation in α-Fe using a hybrid atomistic kinetic Monte Carlo approach. J Chem Phys 135:064502. https://doi.org/10.1063/1.3622045

Castin N, Bakaev A, Bonny G et al (2017) On the onset of void swelling in pure tungsten under neutron irradiation: an object kinetic Monte Carlo approach. J Nucl Mater 493:280–293. https://doi.org/10.1016/j.jnucmat.2017.06.008

Castin N, Bonny G, Bakaev A et al (2018) Object kinetic Monte Carlo model for neutron and ion irradiation in tungsten: impact of transmutation and carbon impurities. J Nucl Mater 500:15–25. https://doi.org/10.1016/j.jnucmat.2017.12.014

Caturla MJ, Aliaga MJ, Martin-Bragado I et al (2016) Microstructure evolution in Fe and Fe-Cr alloys with OKMC methods. EPJ Web Conf 115:03001. https://doi.org/10.1051/epjconf/201611503001

Chang Z, Olsson P, Terentyev D, Sandberg N (2013) Dislocation bias factors in fcc copper derived from atomistic calculations. J Nucl Mater 441:357–363. https://doi.org/10.1016/j.jnucmat.2013.06.029

Chang Z, Terentyev D, Sandberg N et al (2015) Anomalous bias factors of dislocations in bcc iron. J Nucl Mater 461:221–229. https://doi.org/10.1016/j.jnucmat.2015.03.025

Chiapetto M, Malerba L, Becquart CS (2015a) Nanostructure evolution under irradiation in FeMnNi alloys: a "grey alloy" object kinetic Monte Carlo model. J Nucl Mater 462:91–99. https://doi.org/10.1016/j.jnucmat.2015.03.045

Chiapetto M, Malerba L, Becquart CS (2015b) Effect of Cr content on the nanostructural evolution of irradiated ferritic/martensitic alloys: an object kinetic Monte Carlo model. J Nucl Mater 465:326–336. https://doi.org/10.1016/j.jnucmat.2015.06.012

De Backer A, Adjanor G, Domain C et al (2015) Modeling of helium bubble nucleation and growth in austenitic stainless steels using an object kinetic Monte Carlo method. Nucl Instrum Methods Phys Res Sect B 352:107–114. https://doi.org/10.1016/j.nimb.2014.11.110

Domain C, Becquart CS, Malerba L (2004) Simulation of radiation damage in Fe alloys: an object kinetic Monte Carlo approach. J Nucl Mater 335:121–145. https://doi.org/10.1016/j.jnucmat.2004.07.037

Dopico I, Castrillo P, Martin-Bragado I (2015) Quasi-atomistic modeling of the microstructure evolution in binary alloys and its application to the FeCr case. Acta Mater 95:324–334. https://doi.org/10.1016/j.actamat.2015.05.040

Fikar J, Gröger R, Schäublin R (2017a) Interaction of irradiation-induced prismatic dislocation loops with free surfaces in tungsten. Nucl Instrum Methods Phys Res Sect B 393:186–189. https://doi.org/10.1016/j.nimb.2016.10.006

Fikar J, Gröger R, Schäublin R (2017b) Effect of orientation of prismatic dislocation loops on interaction with free surfaces in BCC iron. J Nucl Mater. https://doi.org/10.1016/j.jnucmat.2017.07.032

Garnier T, Nastar M (2013) Coarse-grained kinetic Monte Carlo simulation of diffusion in alloys. Phys Rev B 88. https://doi.org/10.1103/PhysRevB.88.134207

Ghoniem N. Rate theory. Springer. Homer E. Handbook of Materials Modeling, Singapore

Heinisch HL (1983) Defect production in simulated cascades: cascade quenching and short-term annealing. J Nucl Mater 117:46–54. https://doi.org/10.1016/0022-3115(83)90008-9

Heinisch HL, Singh BN (2002) The effects of one-dimensional migration of self-interstitial clusters on the formation of void lattices. J Nucl Mater 307–311:876–880. https://doi.org/10.1016/S0022-3115(02)01005-X

Heinisch HL, Trinkaus H, Singh BN (2007) Kinetic Monte Carlo studies of the reaction kinetics of crystal defects that diffuse one-dimensionally with occasional transverse migration. J Nucl Mater 367–370:332–337. https://doi.org/10.1016/j.jnucmat.2007.03.034

Henkelman G, Jónsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115:9657. https://doi.org/10.1063/1.1415500

Hou J, Kong X-S, Li X-Y et al (2016) Modification on theory of sink strength: an object kinetic Monte Carlo study. Comput Mater Sci 123:148–157. https://doi.org/10.1016/j.commatsci.2016.06.024

Hudson TS, Dudarev SL, Caturla M-J, Sutton AP (2005) Effects of elastic interactions on post-cascade radiation damage evolution in kinetic Monte Carlo simulations. Philos Mag 85:661–675. https://doi.org/10.1080/14786430412331320026

Jansson V, Malerba L (2013) Simulation of the nanostructure evolution under irradiation in Fe–C alloys. J Nucl Mater 443:274–285. https://doi.org/10.1016/j.jnucmat.2013.07.046

Jansson V, Malerba L (2014) OKMC simulations of Fe–C systems under irradiation: sensitivity studies. J Nucl Mater 452:118–124. https://doi.org/10.1016/j.jnucmat.2014.05.011

Jansson V, Malerba L, De Backer A et al (2013) Sink strength calculations of dislocations and loops using OKMC. J Nucl Mater 442:218–226. https://doi.org/10.1016/j.jnucmat.2013.08.052

Jiménez F, Ortiz CJ (2016) A GPU-based parallel object kinetic Monte Carlo algorithm for the evolution of defects in irradiated materials. Comput Mater Sci 113:178–186. https://doi.org/10.1016/j.commatsci.2015.11.011

Johnson MD, Caturla M-J, Díaz de la Rubia T (1998) A kinetic Monte–Carlo study of the effective diffusivity of the silicon self-interstitial in the presence of carbon and boron. J Appl Phys 84:1963–1967. https://doi.org/10.1063/1.368328

Jourdan T, Crocombette J-P (2012) Rate theory cluster dynamics simulations including spatial correlations within displacement cascades. Phys Rev B 86. https://doi.org/10.1103/PhysRevB.86.054113

Jourdan T, Fu CC, Joly L et al (2011) Direct simulation of resistivity recovery experiments in carbon-doped α-iron. Phys Scr 2011:014049. https://doi.org/10.1088/0031-8949/2011/T145/014049

Jourdan T, Bencteux G, Adjanor G (2014) Efficient simulation of kinetics of radiation induced defects: a cluster dynamics approach. J Nucl Mater 444:298–313. https://doi.org/10.1016/j.jnucmat.2013.10.009

Lanore J-M (1974) Simulation de l'evolution des defauts dans un reseau par le methode de monte-carlo. Radiat Eff 22:153–162. https://doi.org/10.1080/10420157408230773

Malerba L, Becquart CS, Domain C (2007) Object kinetic Monte Carlo study of sink strengths. J Nucl Mater 360:159–169. https://doi.org/10.1016/j.jnucmat.2006.10.002

Marian J, Becquart CS, Domain C et al (2017) Recent advances in modeling and simulation of the exposure and response of tungsten to fusion energy conditions. Nucl Fusion 57:092008. https://doi.org/10.1088/1741-4326/aa5e8d

Martin-Bragado I, Rivera A, Valles G et al (2013) MMonCa: an object kinetic Monte Carlo simulator for damage irradiation evolution and defect diffusion. Comput Phys Commun 184:2703–2710. https://doi.org/10.1016/j.cpc.2013.07.011

Martin-Bragado I, Abujas J, Galindo PL, Pizarro J (2015) Synchronous parallel kinetic Monte Carlo: implementation and results for object and lattice approaches. Nucl Instrum Methods Phys Res Sect B 352:27–30. https://doi.org/10.1016/j.nimb.2014.12.081

Martínez E, Marian J, Kalos MH, Perlado JM (2008) Synchronous parallel kinetic Monte Carlo for continuum diffusion-reaction systems. J Comput Phys 227:3804–3823. https://doi.org/10.1016/j.jcp.2007.11.045

Mason DR, Yi X, Kirk MA, Dudarev SL (2014) Elastic trapping of dislocation loops in cascades in ion-irradiated tungsten foils. J Phys Condens Matter 26:375701. https://doi.org/10.1088/0953-8984/26/37/375701

Nandipati G, Setyawan W, Heinisch HL et al (2015) Displacement cascades and defect annealing in tungsten, part II: object kinetic Monte Carlo simulation of tungsten cascade aging. J Nucl Mater 462:338–344. https://doi.org/10.1016/j.jnucmat.2014.09.067

Pannier B (2017) Towards the prediction of microstructure evolution under irradiation of model ferritic alloys with an hybrid AKMC-OKMC approach. PhD dissertation, Université Lille (France)

Rodríguez-Martínez FJ, Castejón-Mochón JF, Castrillo P et al (2017) Kinetic Monte Carlo simulation of phase-precipitation versus instability behavior in short period FeCr superlattices. Nucl Instrum Methods Phys Res Sect B 393:135–139. https://doi.org/10.1016/j.nimb.2016.09.026

Rouchette H, Thuinet L, Legris A et al (2014) Influence of shape anisotropy of self-interstitials on dislocation sink efficiencies in Zr: multiscale modeling. Phys Rev B 90:014104. https://doi.org/10.1103/PhysRevB.90.014104

Schaefer H-E, Maier K, Weller M et al (1977) Vacancy formation in iron investigated by positron annihilation in thermal equilibrium. Scr Metall 11:803–809. https://doi.org/10.1016/0036-9748(77)90079-5

Soisson F, Becquart CS, Castin N et al (2010) Atomistic kinetic Monte Carlo studies of microchemical evolutions driven by diffusion processes under irradiation. J Nucl Mater 406:55–67. https://doi.org/10.1016/j.jnucmat.2010.05.018

Soneda N, Ishino S, Takahashi A, Dohi K (2003) Modeling the microstructural evolution in bcc-Fe during irradiation using kinetic Monte Carlo computer simulation. J Nucl Mater 323:169–180. https://doi.org/10.1016/j.jnucmat.2003.08.021

Souidi A, Hou M, Becquart CS et al (2011) On the correlation between primary damage and long-term nanostructural evolution in iron under irradiation. J Nucl Mater 419:122–133. https://doi.org/10.1016/j.jnucmat.2011.08.049

Stoller RE, Golubov SI, Domain C, Becquart CS (2008) Mean field rate theory and object kinetic Monte Carlo: a comparison of kinetic models. J Nucl Mater 382:77–90. https://doi.org/10.1016/j.jnucmat.2008.08.047

Takahashi A, Sharafat S, Nagasawa K et al (2010) Kinetic Monte Carlo simulation of helium-bubble evolution in ODS steels. J ASTM Int 7:101971. https://doi.org/10.1520/JAI101971

Trochet M, Béland LK, Joly J-F et al (2015) Diffusion of point defects in crystalline silicon using the kinetic activation-relaxation technique method. Phys Rev B 91:224106. https://doi.org/10.1103/PhysRevB.91.224106

Valles G, Panizo-Laiz M, González C et al (2017) Influence of grain boundaries on the radiation-induced defects and hydrogen in nanostructured and coarse-grained tungsten. Acta Mater 122:277–286. https://doi.org/10.1016/j.actamat.2016.10.007

Vattré A, Jourdan T, Ding H et al (2016) Non-random walk diffusion enhances the sink strength of semicoherent interfaces. Nat Commun 7:10424. https://doi.org/10.1038/ncomms10424

Vehanen A, Hautojärvi P, Johansson J et al (1982) Vacancies and carbon impurities in $\alpha$ – iron: electron irradiation. Phys Rev B 25:762–780. https://doi.org/10.1103/PhysRevB.25.762

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3:121–127. https://doi.org/10.1016/0022-3697(57)90059-8

Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Radiation effects in solids. Springer, Dordrecht, pp 1–23

Was GS (2017) Fundamentals of radiation materials science: metals and alloys. Springer, New York

Young WM, Elcock EW (1966) Monte Carlo studies of vacancy migration in binary ordered alloys: I. Proc Phys Soc 89:735. https://doi.org/10.1088/0370-1328/89/3/329

# The Stochastic Nature of Deformation Twinning: Application to HCP Materials

# 59

Irene J. Beyerlein and M. Arul Kumar

## Contents

I. J. Beyerlein (✉)
Mechanical Engineering Department, Materials Department, University of California, Santa
Barbara, CA, USA
e-mail: beyerlein@ucsb.edu; beyerlein@engineering.ucsb.edu

M. Arul Kumar
Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos,
NM, USA
e-mail: marulkr@lanl.gov; marulkr@gmail.com

**Abstract**

Deformation twinning is an important mode of plastic deformation in hexagonal close-packed (HCP) materials. Experimental observations indicate that these twins occur heterogeneously across the polycrystalline microstructure during deformation. The variation is too substantial to be authentically represented by average values, advocating the use of statistical analyses and stochastic models in the studies of HCP material deformation response. This chapter reviews recent efforts to explain the origin of the stochastic nature of twinning and to analyze and simulate deformation twinning in HCP materials from this perspective.

# 1    Introduction

Hexagonal close-packed (HCP) polycrystals deform by two mechanisms: the glide of dislocations and expansion of deformation twins (Beyerlein et al. 2014; Partridge 1967; Yoo 1981; Yoo and Lee 1991). Both dislocations and twins are confined to proceed on particular crystallographic planes and directions in a given crystal. To deform the crystal, dislocations glide on these planes. Twins, on the other hand, shear the atoms on these planes causing them to reposition into a configuration that is crystallographically reoriented from the original crystal. The high-resolution transmission electron microscopy (HR-TEM) image in Fig. 1a shows the twin-parent relationship for $\{10\bar{1}2\}$ tensile twin in HCP Zr and is illustrated in Fig. 1b. Strain is imposed as the twin domain grows in size (Beyerlein and Tome 2010; Capolungo et al. 2009; Kumar et al. 2018). Compared to slip, twins greatly affect the mechanical response (Proust et al. 2007; Salem et al. 2006; Wang et al. 2013a; Wronski et al. 2018), formability and ductility (Barnett 2007a, b; Kumar et al. 2017d), and failure mechanisms (Simkin et al. 2007; Yang et al. 2008; Yin et al. 2008) of HCP metals. Figure 1c presents a typical example of how twinning can affect the mechanical response of an HCP alloy (AZ31 Mg) (Wang et al. 2013a). As shown, many details, such as yield stress, strain hardening, and ultimate strength, are affected by twinning.

Both mechanisms, slip and twinning, occur heterogeneously across the crystal. Heterogeneity in slip is evident at nanometer to micron scales. Dislocations glide in arrays within slip bands that are nanometers in thickness. As deformation proceeds, dislocations can form patterns consisting of dislocation-rich areas separating nearly dislocation-free areas (Agnew et al. 2002; Akhtar and Teghtsoo 1971; Bay et al. 1992; Hughes and Hansen 1997; Kuhlman-wilsdorf and Hansen 1991; Kuhlmann-Wilsdorf 1999; Kumar and Mahesh 2012). Highly resolved microscopy techniques,
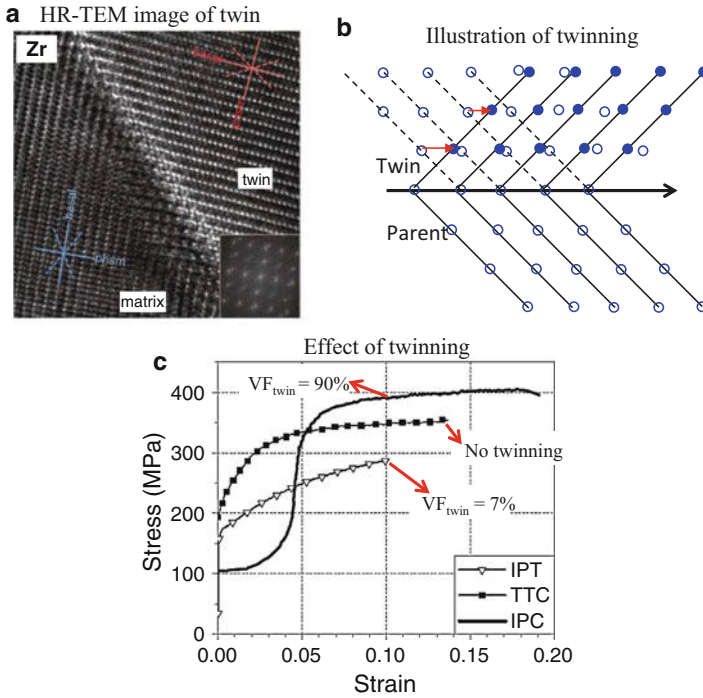
**Fig. 1** (**a**) High-resolution TEM image of the twin-parent orientation relationship for the {1012} twin in HCP Zr (Morrow et al. 2014c). (**b**) Schematic of the shearing of atoms that form the twinned lattice from the parent lattice. (**c**) Example showing the effect of deformation twinning on the stress-strain response of HCP AZ31 Mg alloy (Wang et al. 2013a)

such as TEM, can be used to observe individual dislocations and patterns of dislocations within deformed grains and grain boundaries.

Compared to slip, however, the heterogeneity in twinning is evident at much larger scales, such as those of the polycrystal (Barnett et al. 2012; Beyerlein et al. 2010; Capolungo et al. 2009; Kumar et al. 2018; Wang et al. 2013c). The nonuniform nature of twinning can be easily recognized at the resolution of most standard microscopy and diffraction techniques, such as EBSD and optical microscopy. For instance, for Mg ($c/a = 1.624$), the twin reorientation and local shear associated with the most common tensile twin is 86° and ∼13%, respectively. Large sections of one grain can contain multiple twin domains, whereas another grain of seemingly similar properties (size, shape, orientation) contains no twins. Even in the same grain, the twin thickness can vary easily by two or three times. Twins can expand into neighboring grains. At this larger polycrystal scale, the heterogeneity in dislocation slip would not be discernable.

The widespread heterogeneity in twinning has motivated the use of statistical analyses to understand the role of microstructure, such as grain orientation and grain size, on twinning (Barnett et al. 2012; Beyerlein et al. 2010; Capolungo

et al. 2009; Kumar et al. 2018; Wang et al. 2013c). The apparent sensitivity to local heterogeneities in stress has inspired the use of in-situ testing and simulation techniques to reveal how twin lamella form and expand within a polycrystalline microstructure during deformation (Cheng and Ghosh 2015, 2017a, b; Hazeli et al. 2013; Khosravani et al. 2015; Morrow et al. 2014a, b, c; Wang et al. 2010a, 2011, 2012, 2013b; Wu et al. 2016). Many crystal plasticity models have advanced to incorporate the stochasticity in twin formation and growth, finding better agreement in stress-strain response and microscopy data than the conventional deterministic approach for twin activation (Abdolvand et al. 2015a, b; Abdolvand and Wilkinson 2016; Ardeljan et al. 2015; Beyerlein et al. 2011, 2012; Beyerlein and Tome 2010; Kumar et al. 2017b; Niezgoda et al. 2013, 2014).

The aim of this chapter is to summarize the recent advancements in understanding the formation and growth of deformation twins and likely explanations for the apparent stochastic nature of deformation twinning. The chapter is structured as follows. It begins with a brief review of statistical analyses of large microstructural data sets of materials that twinned during deformation. Next, the modeling methods that have been used to date to simulate twins in HCP polycrystals are briefly introduced. The chapter ends with examples from these calculations and key findings on the role of microstructure and local stresses on twin formation, propagation, and growth.

## 2    Twinning as a Stochastic – Sequential Process

Statistical analysis of twin and stochastic analysis of twinning dynamics can be used as a way toward better understanding of twinning. For many decades, analyses of microscopy data and deformation models have taken a deterministic approach to treating deformation twinning and only recently have they incorporated stochastic aspects, showing noticeable improvements in prediction. The statistical analyses and companion modeling efforts will be discussed in the following sections.

### 2.1    Dynamic Processes of Twinning

The twinning process can be viewed broadly as taking place in three stages. Witnessing these stages in situ is challenging, and in lieu of displaying actual experimental images, a schematic of these stages is provided in Fig. 2, where for simplicity only one twin in one grain is shown. Stage 1 is the creation of an embryo (nucleation), which begins at the atomic scale. Twin nucleation models assume that the twinning partials, needed to create a twin embryo, form directly from the dissociation of pre-existing dislocations. The initiating dislocations may be lattice dislocations (Capolungo and Beyerlein 2008; Cohen and Weertman 1963; Jagannadham 1976; Mahajan and Chin 1973; Mendelson 1972; Priestner and Leslie 1965) or dislocations found in grain boundaries (Barrett and El Kadiri 2014; Beyerlein et al. 2011, 2012; Wang et al. 2014) or moving dislocations that have impinged on bi-phase interfaces (Beyerlein et al. 2013; Zheng et al. 2012). The
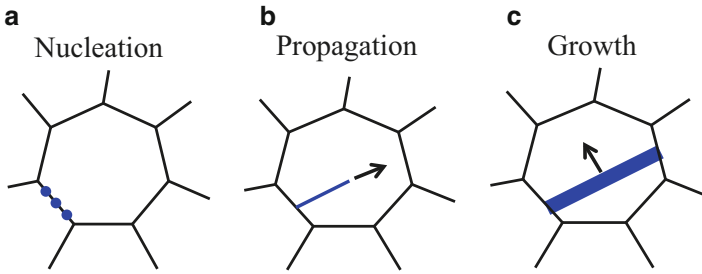
**Fig. 2** Schematic showing three basic sequential steps involved in the formation of twin lamella (Kumar et al. 2015). (**a**) Nucleation, formation of twin nuclei at a given grain boundary; (**b**) propagation, propagation of a twin nucleus into the parent crystal; (**c**) growth, migration of the twin boundary and thickening of the lamella

stresses needed to support these reactions are relatively high, requiring localized stress concentrations, generated for example, at the head of dislocation pile ups.

Stages 2 and 3 involve twin growth and the migration of twin boundaries. In stage 2, growing twin embryos, initially nanoscale to submicron in size, propagate under stress into the grains, often spanning the entire crystal and terminating at the other grain boundaries bounding the same grain (Beyerlein and Tome 2010; Capolungo et al. 2009; Kumar et al. 2018; McCabe et al. 2009). In stage 3, these terminated twins begin to propagate and expand. They can grow either by thickening, so their width increases or by transmitting, a process by which a second twin forms on the other side of the grain boundary where the first twin and grain boundary meet.

Both twin nucleation and expansion can occur only if the combination of required stresses and sufficient density of defects are present. Both aspects, mechanical and material in nature, can vary statistically in space and in time in a deforming microstructure and can result in the statistical appearance of a twin or twins in a grain.

## 2.2  Statistical Features of Deformation Twins

In a polycrystal, grains vary in size, crystallographic orientation, and local grain neighborhood. Most commonly considered microstructural variables that affect the propensity for twinning are grain orientation, grain size, and grain neighborhoods. Using EBSD, twins of thicknesses greater than ∼10 nm can be easily seen. As mentioned earlier, analyses of twins using EBSD data on deformed materials typically find twin lamellae that span the grain completely and are bound by grain boundaries.

Due to the statistical nature of twinning, in order to develop reliable statistical correlations between twin properties and grain microstructure from EBSD data, thousands of grains and twins would need to be analyzed. With the advent of automated EBSD techniques, a twinning microstructure that has formed in thousands of

deformed grains can be assessed relative rapidly facilitating creation of statistically significant data sets. With such data, statistical analyses have been carried out to identify correlations between properties of the twinning microstructure (variant, twin thickness) with the grain microstructure (grain size, grain orientation, and grain neighborhood).

In recent years, statistical analyses of twins have been carried out on a number of HCP metals, Mg, Zr, and Ti, and also Mg alloys (Barnett et al. 2012; Beyerlein et al. 2010; Capolungo et al. 2009; Ghaderi and Barnett 2011; Juan et al. 2015; Kacher and Minor 2014; Khosravani et al. 2015; Kumar et al. 2016a, 2017c, 2018; Shi et al. 2015a, c; Tsai and Chang 2013). The general finding is that while grain size and grain orientation can be strongly correlated with the propensity of twinning, the correlations are not as strong as those expected from a deterministic point of view. For instance, the following are found: (1) not all favorably oriented grains twin, (2) some not favorably oriented grains twin, (3) only 40% of twin variants have highest Schmid factor, (4) not all grains of the same orientation twin, (5) twinned grains contain variable numbers of twins, (6) not all grains of the same size twin, and (7) twins have variable thickness.

To elucidate the statistical variation in twinning grain to grain, the statistical results of $\{10\bar{1}2\}$ tensile twin in HCP Mg and Zr are discussed in the following sections. The materials examined are high-purity polycrystalline Mg and Zr with similar initial textures and, hence, similar grain misorientation distributions (Beyerlein et al. 2010; Capolungo et al. 2009). The Mg has a strong basal texture resulting from rolling, where most of the basal poles are aligned within 30° of the normal direction of the sheet. The Zr has a similar strong basal texture, which was developed via clock rolling rather than conventional rolling (Kaschner et al. 2006). Both materials were compressed at $10^{-3}$/s in an in-plane direction to activate $\{10\bar{1}2\}$ twinning. In order to develop a sufficient number of incipient twins in many grains, Zr was compressed at 77 K to 10% strains, and Mg was compressed at room temperature to 3% strain. Figure 3 shows the sample EBSD images of the twinning microstructure of the deformed Mg and Zr. Using an automated EBSD software, large data sets were generated by analyzing several distinct EBSD scans (Beyerlein et al. 2010; Capolungo et al. 2009). The number of grains and twins investigated totaled 2339 and 8550 for Mg, and 639 and 1065 for Zr.

## 2.3    Statistical Analysis of Crystallographic Orientation Effects

Grains are considered to be well oriented for twinning by a given twin type if they have one or more of the six twin variants bearing a high Schmid factor (e.g., >0.33). From a deterministic viewpoint, the twin variant in a grain most likely to form is the one that has the highest Schmid factor. The Schmid factor (SF) of observable twins is the most common parameter quoted in association with twinning. This reference is especially true when discussing whether twin activation obeys an "Schmid criterion," that is, one that is activated by a resolved shear stress in the twin plane and in the twinning direction (TRSS). The SF is defined as the ratio

**Fig. 3** Sample EBSD images of deformed (**a**) Mg and (**b**) Zr metals showing the activation of tensile twining (Beyerlein et al. 2010; Capolungo et al. 2009). To activate sufficient number of twins, Mg and Zr samples compressed along rolling direction to 3% at RT and to 10% at 77 K, respectively



(a) Pure Mg at 3% compression

TT Direction

100 μm

(b) Pure Zr at 10% compression

20 μm
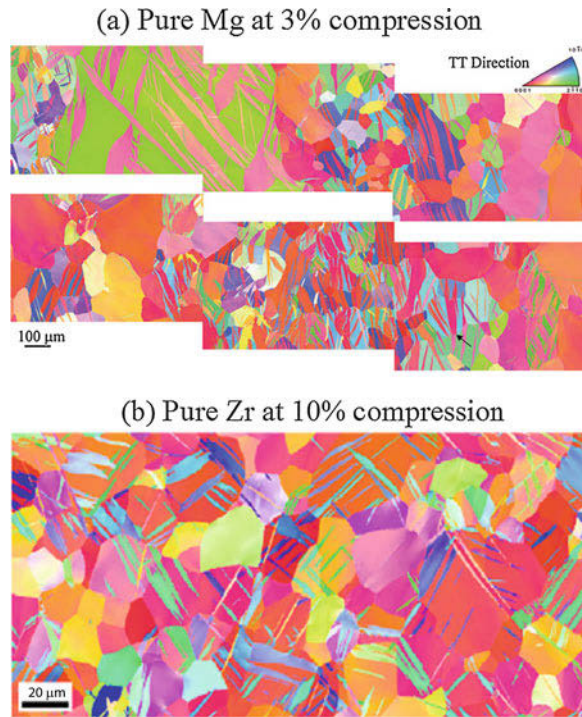
between the TRSS and the value of the macroscopic tensile or compressive stress. The SF varies between $-0.5$ and $0.5$. It provides a geometric measure of how well a twin system is oriented with respect to an external axial stress. Twins belonging to a given grain are classified by their variant and its rank, 1–6, in decreasing order of their SF. If twin activation obeys a Schmid criterion based on the macroscopic applied stress, the twin would correspond to variant 1, the variant having the highest SF among the six in a grain. In this section, the effect of grain orientation as reflected by its SF on twinning formation, growth, and variant selection is discussed.

Figure 4a plots the twinning frequency taken from large data sets on twinned Mg and Zr. The analysis indicates that for both Zr and Mg, the frequency of twinning increases with SF, which is to be expected. However low SF twins are also activated. Figure 4c, d shows the distribution of twin variants as a function of twin SF for Mg and Zr, respectively. The frequency of each twin variant (1–6) is ~35.6%, 23.5%, 21.3%, 11.7%, 3.2%, and 4.6%, respectively, in Mg. Similar frequency in Zr is ~49.8%, 20.0%, 17.9%, 8.2%, 3.3%, and 0.7%. It is surprisingly common, among EBSD studies on twinned microstructures, to find that the twin variant selected is not the one with the highest Schmid factor or even the second highest Schmid factor. These low-rank SF twins have been referred to as "non-Schmid" twins. Activation of non-Schmid twins has also been reported in other HCP metal systems, like HCP Ti and AZ31 Mg alloy (Bieler et al. 2014; Kumar et al. 2018; Shi et al. 2015a, b; Wang et al. 2013c).
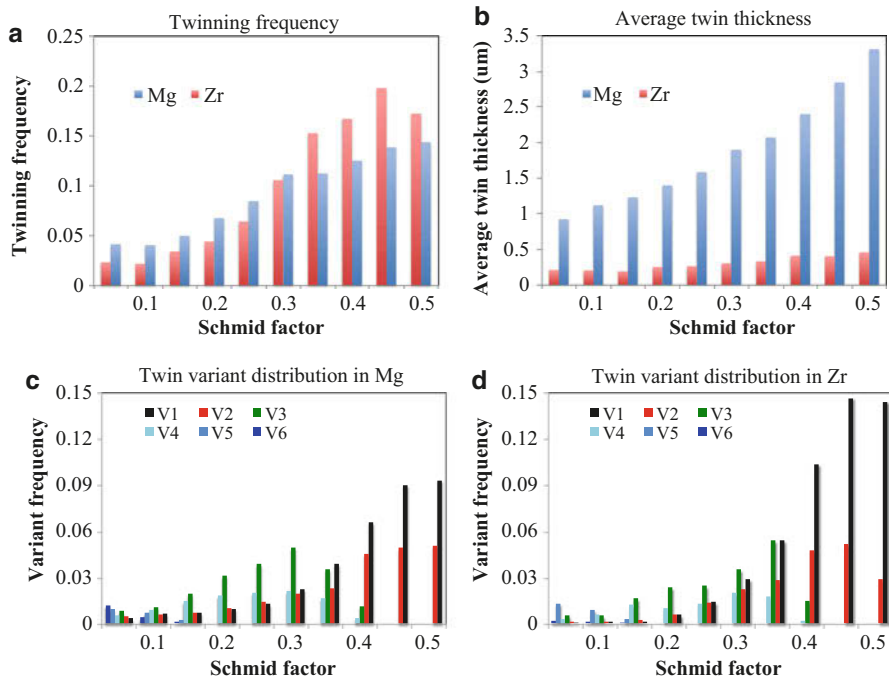
**Fig. 4** Effect of grain orientation, as measured by the Schmid factor associated with the twin system, on (**a**) twinning frequency and (**b**) average twin thickness of $\{10\bar{1}2\}$ tensile twins in deformed Mg (Beyerlein et al. 2010) and deformed Zr (Capolungo et al. 2009). (**c**) and (**d**) are the corresponding twin variant distributions in Mg and Zr, respectively

With all else being the same, twins with high geometric SF are expected to be thicker compared to twins with low SF. From the statistical data set, Fig. 4b shows the distribution of twin thickness as a function of SF for Mg and Zr. Evidently, twins with higher SF are thicker, presumably because they are better oriented for growth.

## 2.4    Statistical Analysis of Grain Size Effects

A common finding in many metals, not only those of HCP crystal structure, is that polycrystals with smaller average grain sizes develop lower twin volume fractions (Barnett et al. 2004, 2012; Beyerlein et al. 2010; Capolungo et al. 2009; Ecob and Ralph 1983; Ghaderi and Barnett 2011; Gutierrez-Urrutia and Raabe 2012; Jain et al. 2008; Juan et al. 2015; Kang et al. 2016; Kumar et al. 2016c, 2018; Lentz et al. 2014; Liu et al. 2015; Rahman et al. 2015; Stanford and Barnett 2008; Tsai and Chang 2013; Wongwiwat and Murr 1978). This frequent observation has motived the application of the Hall-Petch scaling law, originally used for slip, to twinning. This law was first used to explain the higher yield stresses or fracture strains with
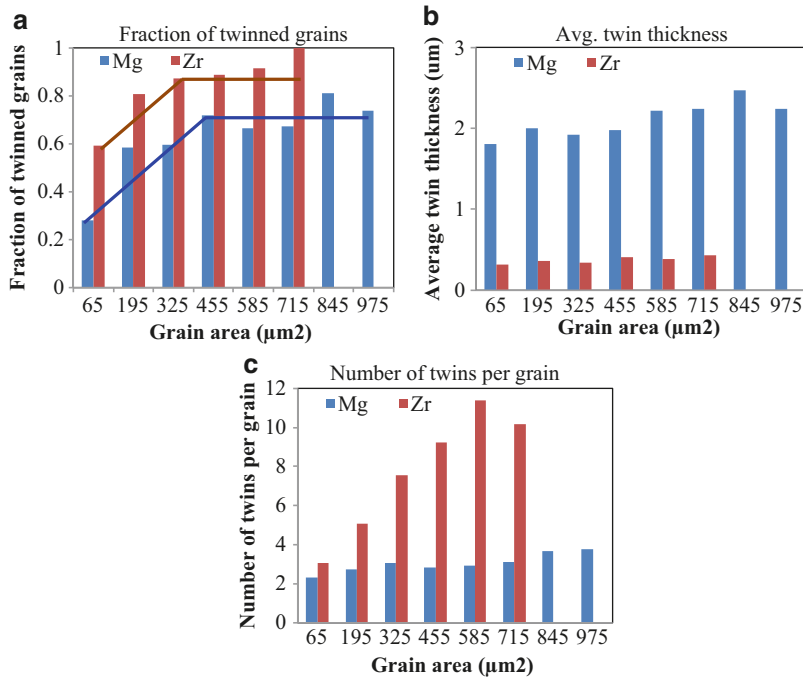
**Fig. 5** Effect of grain size on (**a**) twinning frequency, (**b**) average twin thickness and (**c**) average number of twins per twinned grain of $\{10\bar{1}2\}$ tensile twins in Mg (Beyerlein et al. 2010) and Zr (Capolungo et al. 2009)

decreasing grain size. For twinning, it has been applied in a similar way, such as a twinning stress that increases according $1/\sqrt{D}$, where $D$ is grain size. In nearly all of these cases, the grain size refers to the diameter of the grain that has been cut in the 2D scan. In this section, the grain size dependence on twin frequency from statistically large EBSD data sets are discussed.

The variation in twinning tendency (defined as the number of twinned grains/total number of grains) with grain size for Mg and Zr is shown in Fig. 5a. Here the twinned grains refer to grains with at least one twin of any type. The analysis finds that for both material systems, the propensity for twin activation increases with increasing grain size. It is worth noting that the grain size dependence is not monotonic: the grain size dependence is more pronounced for smaller size grains and tends to saturate for larger grains (see Fig. 5a). The distribution of average twin thickness as a function of grain size is shown in Fig. 5b. The grain size plays a strong role on twin thickness in Mg, but not so in Zr. As another twinning metric, Fig. 5c shows the effect of grain size on the formation of multiple twins in HCP Mg and Zr. In both the materials, the number of twins per grain increases with increasing grain size, but it is particularly striking in Zr. Taking all the data into account, it appears that a grain of a given size in Zr accommodates more twins than grains of the same

size in Mg. Similar twinning statistics has been studied in other HCP metals and for other types of twins (Ghaderi and Barnett 2011; Juan et al. 2015; Lentz et al. 2015, 2016).

## 3 Computational Modeling Methods

### 3.1 Challenges in Modeling the Stochastic Twinning Process

In the foregoing section, statistical aspects of deformation twins were discussed. The substantial spatial variability in these features, across a deformed material grain structure, presents challenges in building both understanding and material models for several reasons. Firstly, twins appear to form randomly as the material is being deformed, and the origins of this stochastic behavior still need to be fully understood. Secondly, the discrete nature of twinning makes it inherently inhomogeneous, making the more commonly used and computationally efficient homogenization mean-field models not readily applicable to treat materials that undergo profuse deformation twinning. Thirdly, twin development is naturally a multiscale process. Twins initiate at the atomic scale, form embryos at the nanoscale, and grow to the submicron scale and can propagate across a grain and grain structure, manifesting at the micron scale and larger. Thus, it is not readily apparent how to apply coarse-graining modeling techniques to deformation twinning. Last, twins do not occur in isolation but concurrently with slip. The shear for HCP metals for the most common $\{10\bar{1}2\}$ tensile twin is 13% for Mg. Thus, even if the entire grain were to twin, slip would need to occur simultaneously in order to plastically strain the material. Twin-slip interactions are just as important or arguably more important for understanding the constitutive response of a material that deforms by slip and twinning. These interactions cannot be fully understood by studying slip and twinning separately.

### 3.2 Some Important Components for Models of Polycrystalline Materials That Deform by Slip and Twinning

Modeling the deformation of polycrystalline materials has been accomplished via a combination of crystal plasticity theory and polycrystal modeling schemes. Crystal plasticity (CP) theory is used to relate the distortion of a strained crystal to slip on crystallographic slip systems (Asaro 1983; Asaro and Lubarda 2006; Hosford 1993). Polycrystal plasticity models then link the individual grain response predicted by CP theory to the overall mechanical response of the polycrystalline aggregate (Asaro 1983; Canova et al. 1988; Kocks 1970; Kocks et al. 2000; Peirce et al. 1982; Roters et al. 2010; Tome et al. 1984). These polycrystal models appear in various levels of sophistication and computational efficiency as will be discussed shortly.

Implementing deformation twinning into a polycrystalline model would ideally seek to include the following elements: (1) the available twin modes, (2) a model

for activating a twin or a number of twins inside a grain, (3) a criterion for variant selection of the activated twins, (4) a scheme for reorienting and shearing the portion of the grain that is twinned, and (5) a criterion to grow the twin. Additional elements include accounting for twin-slip interactions and twin-twin interactions and twin transmission across grains. Not as many models exist that account for these other phenomena, in spite of the fact that they are common occurrences.

## 3.3   Two Categories of Computational Methods for Modeling Twins in Polycrystals

Computational crystal plasticity-based material models for the deformation of poly-crystalline materials can generally be broken down into two categories: mean-field models and full-field, spatially resolved models. The homogenized or mean-field response models, such as self-consistent models, take only a statistical description of the microstructure as input (e.g., crystallographic texture) and return a sample scale or effective response and microstructure evolution (Kocks et al. 2000). In this chapter, the visco-plastic self-consistent (VPSC) model is presented as an example for homogenized models (Lebensohn and Tome 1993).

Full-field micromechanical models, such as crystal plasticity finite element (CPFEM)- or crystal plasticity fast Fourier transform (CPFFT)-based approaches, require as input an explicit spatially resolved description of the material structure (e.g., grain orientation map) and return the spatially resolved material response and local structural evolution (Abdolvand and Daymond 2013b; Abdolvand et al. 2011; Ardeljan et al. 2016; Bronkhorst et al. 1992; Delannay et al. 2006; Eisenlohr et al. 2013; Idiart et al. 2006; Kalidindi 1998; Kanjarla et al. 2012b; Knezevic et al. 2016; Lebensohn 2001; Lebensohn et al. 2008, 2011a, b, 2012; Liu et al. 2010; Masson et al. 2000; Michel et al. 2000, 2001; Mika and Dawson 1999; Moulinec and Suquet 1994, 1998; Shanthraj et al. 2015; Zecevic and Knezevic 2017; Zhao et al. 2007). The term "full-field" indicates that both long-range and short-range grain interactions are considered, and the micromechanical fields are resolved on a discrete grid.

## 3.4   Homogenized VPSC Model Framework

VPSC model describes the polycrystal as a collection of orientations (grains) each with associated volume fraction. Each grain is regarded as a visco-plastic inclusion embedded in, and interacting with, a "homogeneous effective medium" (HEM), which has the average properties of the polycrystalline aggregate. The macroscopic response of the polycrystal results from the contribution of each grain. The visco-plastic compliance of the HEM is given by a self-consistent condition applied on the grain averages. At the single crystal level, the strain rate is given by the individual shear contributions of all active slip and twinning systems in the grain, as:

$$\dot{\varepsilon}_{ij}^g = \sum_s m_{ij}^s \dot{\gamma}^s = \dot{\gamma}_0 \sum_s m_{ij}^s \left( \frac{m_{kl}^s \sigma_{kl}^g}{\tau^s} \right)^n \tag{1}$$

Here, $\boldsymbol{m}^s$ is the symmetric Schmid tensor and $\dot{\gamma}^s$ is the shear rate on system $s$, $\dot{\gamma}_0$ is a normalized shear rate, $n$ is the inverse strain-rate sensitivity, and $\tau^s$ is the threshold or critical resolved shear stress required to activate system $s$. The latter is evolved using a thermally activated dislocation density-based hardening law (Beyerlein and Tome 2008). In order to avoid the further strain-rate dependence associated with the power $n$ in Eq. (1), $\dot{\gamma}_0$ is chosen equal to the norm of the macroscopic strain rate $\left\| \dot{\bar{\varepsilon}}_{ij} \right\|$. The constitutive laws relating strain-rate and stress for a single crystal and for the aggregate are written in a linearized form as:

$$\begin{aligned} \dot{\varepsilon}^g &= M^g : \sigma^g + \dot{\varepsilon}_0^g \\ \dot{\bar{\varepsilon}} &= \overline{M} : \overline{\sigma} + \dot{\bar{\varepsilon}}_0 \end{aligned} \tag{2}$$

where $M^g$ and $\overline{M}$ are the grain and the macroscopic visco-plastic compliance tensors. The tensors $\dot{\varepsilon}_0^g$ and $\dot{\bar{\varepsilon}}_0$ are the back-extrapolated terms for the grain and aggregate, respectively. These variables result from the linearization of Eq. (1). The effect of the linearization scheme on individual grain and polycrystal responses, thus, emerges only through these two variables.

The inclusion formalism couples stress and strain-rate in the grain $(\sigma^g, \dot{\varepsilon}^g)$ with the average stress and strain-rate in the effective medium $(\overline{\sigma}, \dot{\bar{\varepsilon}})$ through an interaction equation:

$$\left( \dot{\varepsilon}^g - \dot{\bar{\varepsilon}} \right) = -\tilde{M} : \left( \sigma^g - \overline{\sigma} \right) \tag{3}$$

where

$$\tilde{M} = n^{\text{eff}} (I - E)^{-1} : E : \overline{M}^{\text{secant}} \tag{4}$$

and $E$ is the visco-plastic Eshelby tensor, $\overline{M}^{\text{secant}}$ is the macroscopic visco-plastic compliance tensor for the secant interaction $(\dot{\bar{\varepsilon}}_0 = 0)$, and the parameter $n^{\text{eff}}$ "tunes" the stiffness of the inclusion-matrix interaction: $n^{\text{eff}} = 0$ for a Taylor case and $n^{\text{eff}} = 1$ for the stiff secant case.

## 3.5    Twinning in SC Approach: CG Model

An aspect of twinning that needs to be incorporated into the models is the fact that twins are finite domains that reorient the lattice and shear portions of the grain (usually taking on a lamellar morphology) and introduce a twin boundary. Over the years, a number of methods have been introduced for treating the
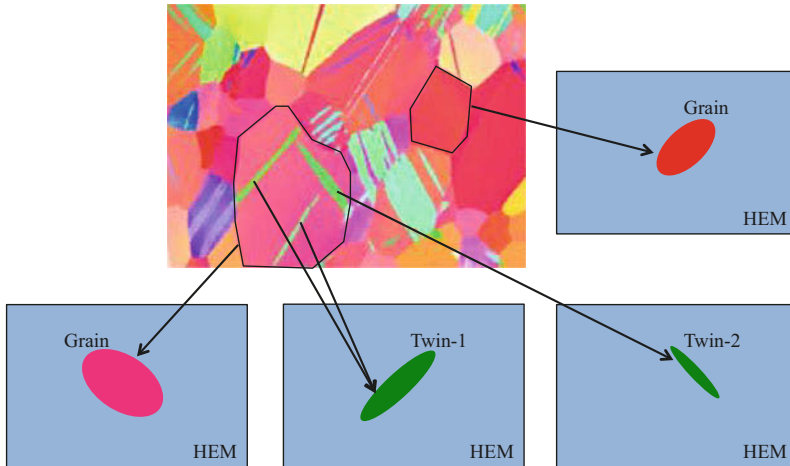
**Fig. 6** Schematic of the composite-grain (CG) model for modeling twin reorientations within the VPSC framework. Un-twinned grain is represented as an ellipsoidal inclusion within a homogenous effective medium (HEM). For twinned grains, which may be comprised of one twin variant type or multiple variants, the matrix and each twin variant is represented as separate inclusions. As an example, "Twin-1" corresponds to two twins of the same variant, and "Twin-2" corresponds to another twin variant, all within the same grain. Here the HEM represents the average response of the polycrystal except the chosen grain and/or twin

reorientation that accompanies twinning: (i) predominant twin reorientation (PTR) method (Lebensohn and Tome 1993; Tome et al. 1991) and composite-grain (CG) method (Proust et al. 2009; Proust et al. 2007), for instance. In these schemes, the twin phase replaces some fraction of the matrix phase, and as the volume of the twin phase increases with strain, the volume of the matrix phase shrinks accordingly. They generally involve splitting the original orientation (grain) into two parts, one part that is twinned and another part that is the matrix, while preserving the original volume fraction of the grain. For instance, in the CG method, the newly twinned grain is split into two inclusions, a twin inclusion with the twinned volume fraction and the remaining parent inclusion with the remaining fraction. The newly formed twin inclusions are treated as new ellipsoidal inclusions and added to the total number of grains in the polycrystal. As grains in these models are represented as ellipsoidal inclusions, the new twin inclusions can be made initially flat ellipsoids with their short axis perpendicular to the twinning plane, to reflect the lamellar shape of newly formed twins (see Fig. 6). The twinned inclusion adopts a mirror orientation with respect to the orientation of the parent grain, that is, characteristic of the type of twin (Yoo 1981; Yoo and Lee 1991).

Later, a modified CG model was developed, which allows for multiple twin types and variants to form in the same grain (Niezgoda et al. 2014). In the modified CG model, the twin and matrix grains are treated as two noninteracting inclusions (grains) embedded in the homogeneous effective medium. It is schematically shown

in Fig. 6. In the figure, a grain with two types of twin variants is shown, and in this case, un-twinned matrix grain, twin of variant 1, and twin of variant 2 are considered separate inclusions in the effective medium. The two twins are initially given a flat ellipsoid shape. The shortest axis of the ellipsoid is parallel to the twinning plane normal and another of the ellipsoid axes is parallel to the twinning direction. A similarly oriented ellipsoid is created to represent the un-twinned region of the grain. The aspect ratios of both ellipsoids evolve with deformation. These twin and matrix grains are characterized by independent secant compliances $M^{\mathrm{sec}}$, and consequently no explicit twin-matrix interaction is considered when solving the self-consistent equations. The relative fraction of each phase is updated incrementally with deformation as the grain twins.

Once a twin has formed inside a grain, growth of this twin is determined in a more traditional deterministic fashion within the modified CG framework. During deformation, the twin shear rate $\dot{\gamma}^{v}$ is calculated for the nucleated twin or variant $v$ in each grain by:

$$\dot{\gamma}^{v} = \dot{\gamma}_0 \left( \frac{\tau^{v}}{\tau_{c}^{\mathrm{twin}}} \right)^{n} \tag{5}$$

Here $\tau^{v}$ and $\tau_{c}^{\mathrm{twin}}$ are the resolved shear stress on twin variant v, and critical resolved shear stress for twin domain expansion (after nucleation).

## 3.6    Full-Field CPFFT Model Framework

The FFT-based crystal plasticity-based models provide spatially resolved micromechanical fields in the individual crystals within polycrystals. The formulation provides an exact solution of the governing equations of equilibrium and compatibility, in such a way that the final (converged) equilibrated stress and compatible strain fields fulfill the constitutive relationship at every discrete material point.

The original FFT formulation was developed to study the local and effective mechanical response of linear elastic (Moulinec and Suquet 1994), nonlinear elastoplastic (Moulinec and Suquet 1994, 1998), and visco-plastic (Michel et al. 2000, 2001) composite materials. The FFT formulation was later adapted for polycrystalline materials and permitted the study of the effective and local responses associated with the heterogeneity in the spatial distribution of crystallography and directional dependence of mechanical properties (Lebensohn 2001). In recent years, this FFT formulation has been extended to different deformation regimes like elasticity (Brenner et al. 2009), incompressible visco-plasticity (Lebensohn 2001; Lebensohn et al. 2008), dilatational visco-plasticity (Lebensohn et al. 2011b), infinitesimal elasto-visco-plasticity (Kanjarla et al. 2012b; Lebensohn et al. 2012) and finite elasto-visco-plasticity (Eisenlohr et al. 2013). Below the CPFFT model that allows for the crystals to deform by infinitesimal elasto-visco-plasticity (EVP) is briefly described and applied in later examples to study stress fluctuations in a deformed polycrystal and local stress states generated around discrete twins.

In the CPFFT method, the solution of an EVP problem involves the adoption of an appropriate time discretization scheme. Using an Euler implicit time discretization and Hooke's law, the expression for the stress in material point $x$ at $t + \Delta t$ is given by:

$$\boldsymbol{\sigma}^{t+\Delta t}(x) = \mathbf{C}(x) : \boldsymbol{\varepsilon}^{e,t+\Delta t}(x) = \mathbf{C}(x) : \left(\boldsymbol{\varepsilon}^{t+\Delta t}(x) - \boldsymbol{\varepsilon}^{p,t}(x) - \dot{\boldsymbol{\varepsilon}}^{p,t+\Delta t}(x)\,\Delta t\right)$$

(6)

where $\sigma(x)$ is the Cauchy stress tensor; $\mathbf{C}(x)$ is the elastic stiffness tensor; $\varepsilon(x)$, $\varepsilon^e(x)$, and $\varepsilon^p(x)$ are the total, elastic, and plastic strain tensors; and $\dot{\varepsilon}^p(x)$ is the plastic strain-rate tensor given in Eq. (1). The CPFFT model solves the equilibrium equation along with the above constitutive equation by iteratively adjusting the compatible strain field at every material point. The spatially resolved local response is calculated using the convolution integral between the Green's function associated with the displacement field of a linear reference homogeneous medium and a polarization field in which the heterogeneity and nonlinearity of the problem is specified. Application of Fourier transforms reduces the convolution integrals of the equilibrium equation in real space into simple products in Fourier space. Specifically the FFT algorithm transforms the polarization fields of the periodic microstructures, which are functions of the unknown strain field, into Fourier space, to obtain the micromechanical responses in real space.

## 3.7 Twinning Model in CPFFT Framework

Up to now, advancing CP models for discrete twins with a 3D microstructure has been hindered by the lack of 3D microstructural representation codes. Very recently, a few full-field, spatially resolved polycrystal models, such as CPFFT and CPFEM, have been advanced to include discrete twin domains within individual grains (Abdolvand and Daymond 2013a, b; Abdolvand et al. 2011, 2015b; Abdolvand and Wilkinson 2016; Ardeljan et al. 2015). To model the twin domain, the boundary of the domain, the twinning orientation, and the characteristic twin shear needs to be imposed homogeneously throughout the domain. A few examples of the local stress fields calculated around twins in Zr and uranium (U) with these techniques are shown in Fig. 7 (Abdolvand and Wilkinson 2016; Ardeljan et al. 2015). An important aspect captured is the heterogeneous stress field within crystals that result from the twin. The character, intensity, and extent of these fields depend sensitively on the elastic and plastic properties of the material. Recently a series of studies were undertaken to understand how response of the surrounding crystal could impact the mesoscopic processes of twinning (Abdolvand et al. 2018; Kumar et al. 2015, 2016a, b, c, 2017b).

In this section, the recent extension of the CPFFT formulation to account for the reorientation and twinning shear transformation in discrete regions within a crystal is described. In this model, deformation twinning is treated as a shear transformation
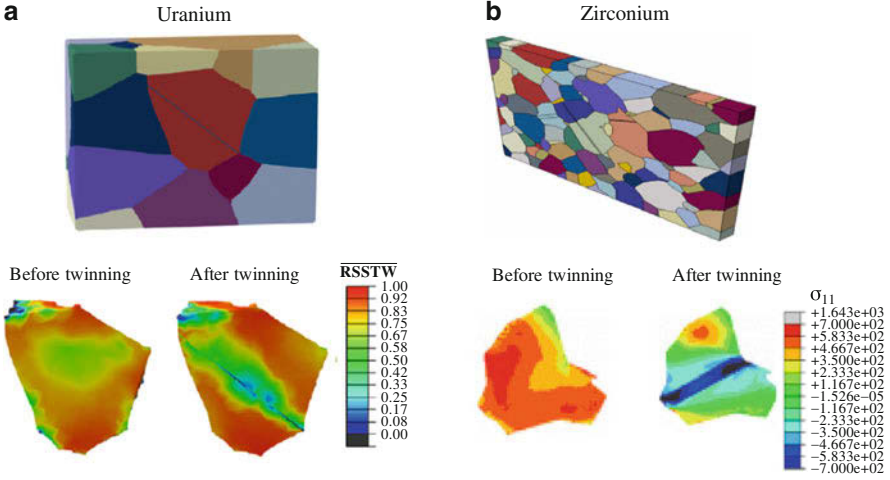
**Fig. 7** Examples of recent modeling efforts in which discrete twin lamellae have been included within a full-field spatially resolved technique. The two shown are crystal plasticity finite element (CPFEM) models for (**a**) uranium (Ardeljan et al. 2015) and (**b**) zirconium (Abdolvand and Wilkinson 2016)

process. Accordingly the constitutive behavior of an elastic-visco-plastic material under an infinitesimal strain approximation with shear transformation becomes:

$$\boldsymbol{\sigma}^{t+\Delta t}(x) = \mathbf{C}(x) : \left(\boldsymbol{\varepsilon}^{t+\Delta t}(x) - \boldsymbol{\varepsilon}^{p,t}(x) - \dot{\boldsymbol{\varepsilon}}^{p,t+\Delta t}(x)\,\Delta t - \boldsymbol{\varepsilon}^{tr,t}(x) - \Delta\boldsymbol{\varepsilon}^{tr,t+\Delta t}(x)\right) \tag{7}$$

where $\boldsymbol{\varepsilon}^{tr}$ is the transformation strain. During the buildup of the twinning transformation, successive shear increments are imposed in the twin domain and the system relaxed. The associated strain increments have the following relationship with the local twin variant at point x:

$$\Delta\boldsymbol{\varepsilon}^{tr}(x) = \mathbf{m}^{tw}(x)\,\Delta\gamma^{tw}(x) \tag{8}$$

For material points lying outside the twin domain, $\Delta\varepsilon^{tr}(x)$ is zero. The tensor $\mathbf{m}^{tw} = \frac{1}{2}\left(\mathbf{b}^{tw} \otimes \mathbf{n}^{tw} + \mathbf{n}^{tw} \otimes \mathbf{b}^{tw}\right)$ is the Schmid tensor associated with the twinning system, where $\mathbf{b}^{tw}$ and $\mathbf{n}^{tw}$ are unit vectors along the twinning direction and twin plane normal, respectively. The twinning transformation builds up in increments, until reaching the characteristic twin shear, $s^{tw}$:

$$\Delta\gamma^{tw}(x) = \frac{s^{tw}}{N^{twincr}} \tag{9}$$

The time increment $\Delta t$ and the number of increments to achieve the twin transformation $N^{twincr}$ are set sufficiently low and high, respectively, to ensure convergence.

# 4       Stochastic Twin Nucleation Model

## 4.1      Probability Model for Critical Stresses for Twin Formation

In a few recent works (Beyerlein et al. 2011; Beyerlein and Tome 2010; Niezgoda et al. 2014), an approach was developed to include grain boundary-induced twin nucleation into constitutive laws for application into the mean-field crystal plasticity models, like VPSC with a twin reorientation scheme. Unlike conventional polycrystal models to date, the model they used for twin nucleation is not deterministic but probabilistic, dictated by the likelihood of forming a twin embryo in the grain boundaries. The approach involves incorporating two aspects of the grain boundaries in a bulk average probabilistic sense. One aspect is a probability model for the nucleation of twins when some numbers of grain boundary defects undergo stress-driven transformations, which then coalesce into a single stable nucleus. It assumes that the time scale of the transformation and subsequent coalescence is instantaneous compared to the applied deformation and introduces a characteristic length scale, within which a critical number of transformations occur to produce a propagating twin. The stochastic model gives an explicit form for the probability distribution for the critical stress values required for twin nucleation that could be used in the VPSC model for activating twinning.

The other aspect concerns the stresses that activate twinning. These stresses are those that are generated at grain boundaries, and these tend to deviate significantly from the average stresses calculated in VPSC for each grain. To tackle this, distributions of grain boundary stresses were obtained from separate full-field CP calculations. Taken together the VPSC model simulations of deformation were advanced to activate twinning when randomly sampled critical twin stresses were exceeded by randomly sampled grain boundary stresses.

In this section, the model is briefly reviewed. Consider a grain within a polycrystal as illustrated in Fig. 8. It has $n_f$ nearest neighbors and is joined to each neighbor k (k = 1, ... $n_f$) by a common grain boundary facet of area $A_k$. Connecting these facets is a network of triple lines and quadruple points, which generally have an atomic structure distinct from those of the facets.

A grain boundary area $A_k$ contains defects, or grain boundary dislocations (GBDs), varying spatially and temporally in defect content (e.g., size of the Burgers vector). When provided a sufficiently high stress for a GBD of a given size, GBDs can transform into a twin embryo, a process observed in atomistic simulation to involve the reshuffling of the atoms into the twinned structure (Wang et al. 2013b). Neighboring smaller twin nuclei can coalesce into a larger twin embryo. If the embryo reaches a critical size, it will propagate into the crystal (Beyerlein and Tome 2010).

According to this physical picture, forming an embryonic twin relies on the right stress fluctuation simultaneously hitting the right-sized defect. This confluence of events is more likely to occur in the grain boundaries, where local stress states and defect content are high. It is usually the case that the spatial distribution of GBDs and stresses are heterogeneous, a GBD-to-twin nucleus transformation is a

statistically occurring event, and the number of such events $N$ is a random variable. The number $N$ is expected to increase as the area $A_k$ of the facet and the magnitude of the stresses acting on the boundary increase.

In the model, a stochastic counting process is proposed for $N$, i.e., $\{N(A_k), A_k \geq 0\}$, where $N(A_k)$ is the number of transformation events that occurred in an area $A_k$. If each event is independent and identically distributed (i.i.d.), and the number of events in nonoverlapping elements is independent (stationary increments), then the Poisson process emerges as an appropriate model. Accordingly, the probability that $N = m$ defects will be transformed into a twin nucleus within a given area $A_k = a$ follows a Poisson distribution:

$$P\,(N = m, a) = \frac{(\lambda a)^m}{m!} \exp\,(-\lambda a) \tag{10}$$

where $\lambda$ is the rate of the process. The parameter $\lambda$ also corresponds to the expected number of transformation events per unit area. Another consequence of the Poisson model is that the events are uniformly distributed over area $a$, and spatial separation between transformation events is exponentially distributed. As the process is driven by stress, the Poisson rate $\lambda$ is assumed a monotonically increasing function of the resolved shear stress ($\tau$) on a twin system $s$. For convenience, the following power law formulation is introduced:

$$\lambda\,(\tau) = \frac{1}{a_0} \left( \frac{\tau}{\tau_0} \right)^{\alpha} \tag{11}$$

where $a_0$ is a material parameter assumed to be constant, and $\tau_0$ is a characteristic scalar stress value, which is interpreted as the stress required to dissociate, on average, one grain boundary dislocation on area $a_0$. Parameters $a_0$ and $\tau_0$ are, in

principle, functions of the defect content of the grain boundary. Therefore, they may vary from boundary to boundary and differ, for instance, for a coherent boundary versus an incoherent one.

To implement the above model into a computational mechanics code, the discrete counting of the number of transformation events needs to be linked to a continuous probability of forming a twin nucleus. To this end, a characteristic area $a_c$ is introduced as the minimum area that can produce one twin, i.e., one characteristic area $a_c$ produces one crystalline twin, which may have resulted from the coalescence of $n$ tinier twin nuclei created by $n$ transformation events. Suppose that at least $m^*$ distinct conversion events need to occur within $a_c$ in order to form one twin, then the probability that at least $m^*$ events occur in $a_c$ is:

$$P\big(N \geq m^*, a_c\big) = 1 - \sum_{m=1}^{m*-1} P(N = m, a_c) \tag{12}$$

Further, we assume that $m^* = 1$, that is, at least one defect must be activated within $a_c$, which yields the following Weibull distribution.

$$P(S < \tau) = P(N \geq 1, a_c) = 1 - \exp\left(-\left(\frac{\tau}{\tau_c}\right)^{\alpha}\right) \tag{13}$$

Using Eqs. (12) and (13), we can redefine P($S < \tau$) as the probability that the critical stress to nucleate a twin is less than or equal to $\tau$. Here $S$ is a random variable that quantifies the critical nucleation strength or equivalently the stress required to transform an appropriate number of grain boundary dislocations into twinning dislocations.

The area $a_c$ is an important model length scale, where in all dissociation events in $a_c$ lead to one twin. Accordingly, it sets the minimum twin spacing and the maximum number of twin lamellae that can form from a grain boundary of area $A_k$ as $n^* = A_k/a_c$. The material parameter $\alpha$ governs the dispersion in $S$ and is linked to the type of defects in the grain boundary.

## 4.2    Statistical Representations of Grain Structures

The above model applies to twinning from one-grain boundary facet of area $A_k$. Multiple grain neighbors in fact surround any given grain, and each neighbor shares just one of the many facets that comprise a grain's boundary.

Recall from Sect. 4.1 that in a given grain facet $A_k$, there are $n^*$ number of potential sites to form a twin. Correspondingly, following Eq. (13), there is a set of $n^*$ critical stresses $S_i$, $i = 1, \ldots n^*$, assumed to be independent and identically distributed. In order to assign the number of potential nucleation sites $n^*$ for each grain, the three-dimensional grain structure is needed. At a minimum, the microstructural parameters needed are (1) the number of neighbors ($n_k$) or

number of grain boundary facets of a grain and (2) the surface area of the grain boundary facets $A_k$ from which the twins will nucleate. These parameters depend on the grain size distribution and the morphology of the grains in the sample and can in principle be estimated from a combination of microscopy and some basic stereological principles.

When the computational model is a mean-field model, like VPSC, every grain is modeled as an ellipsoidal inclusion. Thus estimates are needed for the lattice orientation, effective radius (or equivalently volume), and the ratio between major and minor axes of the grains. For this purpose, the stochastic field model of Thorvaldsen (1993) is adopted. Using simple geometric arguments, this model develops an expression for the distribution of neighbors for a given grain and the distribution of facet areas depending on its size, $R$, and shape. According to this model, for a spherical grain with radius $R$, the expected number of grain neighbors $n_f$, $E(n_f)$, is related to the expected value of $R$, $E(R)$ according to:

$$E\left(n_f\right) = 4\left(1 + \frac{R}{E(R)}\right)^2 \tag{14}$$

Given a neighboring grain with radius $R_n$, the contact area, $A_k$, is given by (Thorvaldsen 1993):

$$\frac{1}{A_k} = \frac{1}{\pi}\left(\frac{1}{R_n} + \frac{1}{R}\right)^2 \tag{15}$$

To demonstrate, the mode is applied to high-purity HCP Zr. Figure 9 shows the distribution of grain size (Capolungo et al. 2009), the predicted distribution for the number of neighboring grains or facets, and the predicted distribution of grain facet areas for Zr. The equivalent circle radius exhibits an approximate Rayleigh distribution. Accordingly, the cross-sectional grain areas will be exponentially distributed. The expected number of neighbors for this grain size distribution is 17, indicating a wide range of grain sizes in the Zr aggregate being modeled.

## 4.3 Spatially Resolved Stress States in Deforming 3D Polycrystals

This critical stress $\tau_c$ can then be compared to the twin-plane resolved shear stress (TRSS) for the twin variant in question. If it exceeds the TRSS, then a twin embryo may nucleate, and if not, then the region remains un-twinned. For twinning, the local stress state in the region in question for twinning is desired. Self-consistent schemes can be used to calculate the local stress state in the grain. However the formulation does not allow for calculation of the stress state in the boundaries. Also by modeling each grain as an ellipsoid in a homogeneous medium, the scheme does not provide a way of defining the orientation of the grain boundary with respect to the loading direction. Full-field, spatially resolved techniques such as CPFFT or CPFE, however, are particularly well suited for calculating stress distributions
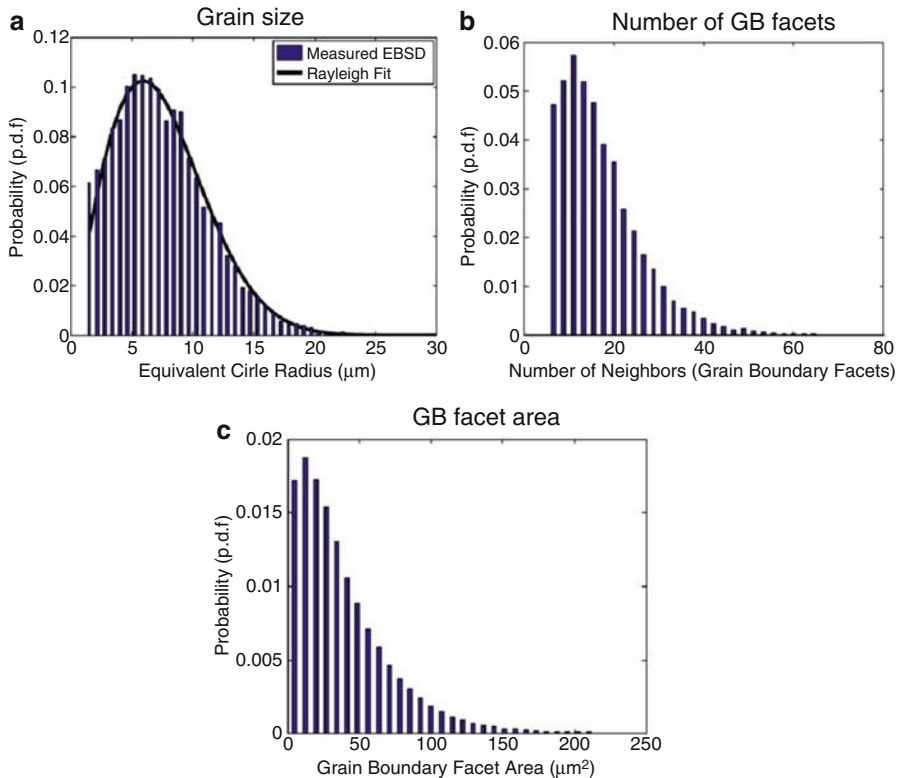
**Fig. 9** Microstructural variations considered in the stochastic twin nucleation model: (**a**) grain size, (**b**) number of grain boundary facets, and (**c**) grain boundary facet area (Niezgoda et al. 2014). These microstructural distributions are obtained from the EBSD-based statistical analysis of pure Zr (Capolungo et al. 2009)

within grains and in grain boundaries and triple and quadruple points. From this model, the stress fluctuation distributions, the deviation in stress from the average value, can be calculated and amenable for use in a mean-field technique.

Following this approach, CPFFT calculations are carried out on Zr. Several (100) realizations of 3D grain structures of representative Zr material volumes were performed to extract the stress fluctuations at grain boundaries (Kanjarla et al. 2012a; Niezgoda et al. 2013). The three-dimensional (3D) representative volume elements, RVEs, were created using DREAM-3D. Each RVE contained $500 \pm 50$ grains with $\sim 5 \times 10^5$ grain boundary voxels or elements per RVE. For each representative volume, simulations of multiple loading conditions were performed. The deviations from the grain-averaged stress at the grain boundaries were extracted for all representative volumes for each loading condition.

For one example RVE, Fig. 10a shows calculated stress fluctuations, defined as deviations in the Von-Mises effective stress from the grain average stress, at 2.5%
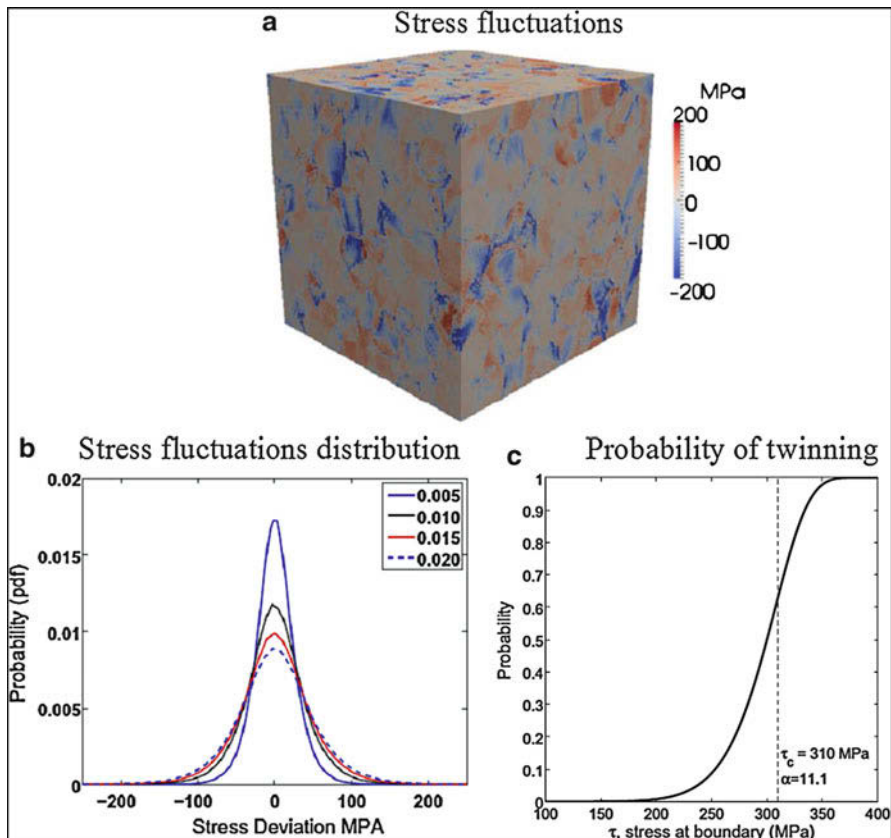
**Fig. 10** Mechanical stress fluctuations as calculated in the CPFFT model: (**a**) the deviations in the effective stress from the grain average stress. The FFT model contains 522 grains (Niezgoda et al. 2013). (**b**) Distribution of the CPFFT calculated stress fluctuations from 100 different representative volume elements like those in (**a**) (Niezgoda et al. 2013). (**c**) Separately, the probabilistic nucleation model provides the probability distribution for the critical stress to form a twin (Niezgoda et al. 2014)

applied strain. This result reveals significant deviations at the grain boundaries. It is not uncommon for single grains to have regions near their boundaries that deviate in both a highly compressive and tensile state from the grain average. Figure 10b presents the observed deviations in the normal stress components from the grain stress. By averaging over multiple loading conditions, the fluctuations on all three normal components had approximately the same distribution, as did the three shear stress components. As shown, the distributions are approximately Gaussian; however, the extreme tails of the distributions extend significantly farther than would be expected from a perfect Gaussian. For VPSC deformation simulations, a six-dimensional Gaussian with a zero mean vector was used to reasonably represent the CPFFT calculations.

## 4.4     Incorporation of Statistical Stress and Strength Distributions in Homogenization Models

Carrying out the simulation with the probability model for twinning added, the VPSC model needs as input the distribution for the twin nucleation stresses and stress fluctuations. Figure 10c shows the former, the twin nucleation stress distribution for $a_c = 0.01$ μm$^2$, $\alpha = 11.1$, and $\tau_c = 310$ MPa. The latter stress fluctuations, just described in the prior section, are provided in Fig. 10b. In an actual polycrystal, grain boundary misorientations are distributed, and their defect states vary, and in deformation, both can vary in time. However, for simplicity, the same Gaussian representation for the stress fluctuation distribution is applied for all grain boundaries at all times during deformation.

The procedure used is described as follows. The calculation begins with a description of the microstructure. The input texture for VPSC is constructed by sampling 8500 grains from the EBSD maps produced by Capolungo et al. (2009). The equivalent circle diameter from the EBSD is used to assign the weight or grain size for each grain (Fig. 9a). For each grain, Eq. (14) is used to determine the number of neighbors. The neighbors are randomly sampled from the input texture, and Eq. (15) is used to assign facet areas $A_k$. Each facet is then divided into $n^*$ areas of size $a_c$. Each of these areas is a potential nucleation site for a twin of variant $v$. For each site, at each strain increment, Eq. (13) is used to test for the nucleation of twin variant $v$, which requires knowing the resolved shear stress $\tau(v)$ projected on the twin variant $v$. To calculate $\tau(v)$, a stress fluctuation $\Delta\sigma$ is randomly sampled from the characteristic distribution (Fig. 10) for each variant and is added to the grain stress calculated at that strain increment. The sum is then projected onto the twin system to compute $\tau(v)$. This procedure for nucleation is then repeated for every twin variant at each site. In the case that multiple twin variants could nucleate from the same site, a single variant is selected at random to propagate. Each nucleation site is tested independently for nucleation, with no correlation with the neighboring sites. Although nucleation could occur on more than one facet belonging to a grain, in these example calculations, one facet for each grain is randomly chosen as the one from which twins are allowed to grow. Using the above procedure, any variant could nucleate, not just the one with the highest $\tau(v)$. Once a grain forms twins, no further nucleation is allowed. Twin growth follows a deterministic stress-based criterion, Eq. (5).

Using the above model, the deformation response of HCP Zr at three different temperatures, liquid nitrogen (77 K), 150 K, and room temperature (300 K) is simulated. In this calculation plastic deformation of Zr is accommodated by prismatic <a> slip and pyramidal <c + a> slip and $\{10\bar{1}2\}$ tensile and $\{11\bar{2}2\}$ compression twins. The evolution of the critical resolved shear stresses for only the slip modes follows the dislocation density-based hardening model developed in Beyerlein and Tome (2008). The same twin nucleation stress distribution is used at all temperatures, while the effect of temperature is included in the initial CRSS and the dislocation density evolution rate for slip. Figure 11a compares the model-predicted stress-strain curves at 76 K, 150 K, and 300 K under in-plane
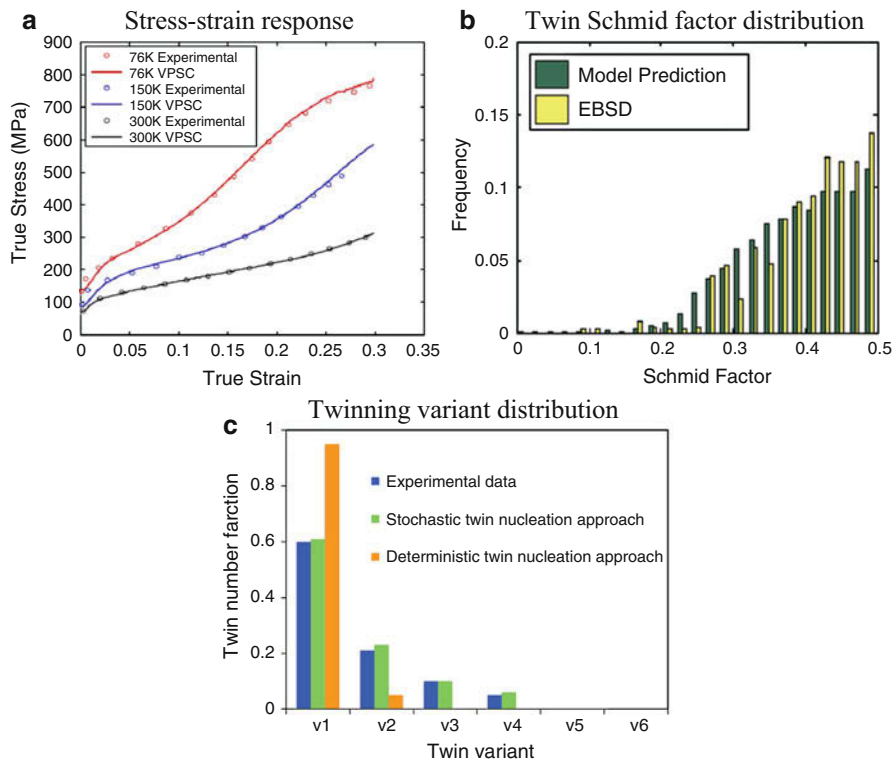
**Fig. 11** Comparison of results from the VPSC model, with the twin nucleation model incorporated, for the (**a**) stress-strain response, (**b**) twinning frequency as a function of Schmid factor, and (**c**) twin number fraction for different twin variants at 10% compression. In (**c**), the frequency of twin variants formed as predicted from the polycrystal VPSC model with and without the stochastic twin model are compared. Without it, the criterion for forming a twin is deterministic. As shown, better agreement with the EBSD data is achieved with stochastic twin formation implemented (Niezgoda et al. 2014)

compression with the experimental curves for Zr. As shown, good agreement in the calculated flow stresses and hardening rates for all temperatures tested is achieved. Although not shown, the model also predicts well the experimentally observed texture evolution (Fig. 6 of Niezgoda et al. 2014)) and twin volume fraction evolution (Fig. 8 of (Niezgoda et al. 2014)). For completeness, the calculated twin volume fraction at 5% and 10% compression for 76 K temperature loading is 4% and 16%, respectively, which agrees well with the measured fraction from EBSD images is 5% and 16%, respectively.

The importance of the stochastic twin nucleation model cannot be fully appreciated from analyzing average responses. The model results can also be compared to local, microstructure data distributions of twinning. Figure 11b compares the calculated number fraction of twins with a given Schmid factor after 10% compression at

76 K with the experimentally measured number distribution of the observed twins at the same strain from EBSD. Like the measurement, the calculated frequency is also broad, including twins with high Schmid factors and with low Schmid factors (m < 0.3). Figure 11c compares the model predicted frequency distribution of each twin variants with that from EBSD. The model finds that the twin variant selected is most likely (but not always) the one with the highest geometric Schmid factor and that the grains that twin are most often (but not always) well oriented for twinning, again, in agreement with the measurement. EBSD analysis indicates that only 50–60% of the twins correspond to the twin variant with the highest SF, i.e., v-1, and 20% to the second highest SF variant, v-2. Remaining ∼20% of twins are the third and fourth highest SF variants.

The model overall predicts well the macroscopic deformation response because it predicts at the mesoscopic scale the formation of a broad range of twin variants over a broad straining period. In contrast, the model employing a deterministic twin nucleation approach, using a constant CRSS for all twins, predicts that ∼95% of the twins activated are v-1 and the remaining ∼5% are v-2. Lower ranked SF twin variants are not activated. The combined analysis of experiment and model results provides evidence that twin formation in polycrystalline materials is stochastic.

## 5   Stochastic Twin Growth Model

The earlier sections in this chapter discussed twin formation, as being derived from reactions involving intense localized stress and dissociations of individual or groups of discrete dislocations. The remaining sections of this chapter focus on the thickening of the twin band. In this stage, the dimensions of the twin domain lie above the atomic scale, being approximately submicron or micron and closer to the dimensions of the parent grain. The stress fields surrounding twin domains of this size range develop as a result of elastic and plastic deformation. The latter is carried by the collective glide of dislocations and can be adequately modeled at this scale by crystallographic slip. Accordingly, to study the effects of slip and crystal orientation on these two stages of deformation twinning, we employ a full-field, spatially resolved technique, the CPFFT model, which permits calculation of the mechanical fields (stress, strain, strain rate) in the presence of discrete twin domains within crystals. In the examples that follow, the model is applied to $\{10\bar{1}2\}$ tensile twins in HCP Mg and Zr metals. The anisotropic elastic constants (in GPa) for Mg at room temperature are 59.75, 23.24, 21.70, 61.70, and 16.39 (Hearmon 1946; Simmons and Wang 1971) and for Zr at 77 K are 143.50, 72.50, 65.40, 164.90, and 32.10 (Fisher and Renken 1964; Simmons and Wang 1971). Plastic deformation is accommodated by basal <a>, prismatic <a>, and pyramidal <c + a> slip for both HCP Mg and Zr metals. The critical resolved shear stress (CRSS) for basal, prismatic, and pyramidal slip systems (in MPa) are 3.3, 35.7, and 86.2, respectively, for Mg at room temperature (Beyerlein et al. 2011), and 700.0, 20.0, and 160.0, respectively, for Zr at 77 K (Knezevic et al. 2015).

## 5.1    Local Twin Boundary Stresses to Expand Twin

As mentioned earlier, twins most often nucleate at a grain boundary and propagate across the grain until they are stopped at the opposing grain boundary. They thicken further after having terminated at grain boundaries. Both the surrounding parent and neighboring grains deform in order to accommodate the shear strains resulting from the growing twin. To investigate the driving forces for twin growth, it is important to understand the local stresses generated around the twin band. To this end, consider a tri-crystal of three neighboring grains embedded in a polycrystal, as shown in Fig. 12. The central grain contains the deformation twin, which spans the width and intersects with the two neighboring crystals. The parent grain orientation is (0°,0°,0°), (Bunge convention), which corresponds to alignment of its c-axis with the Z-direction.
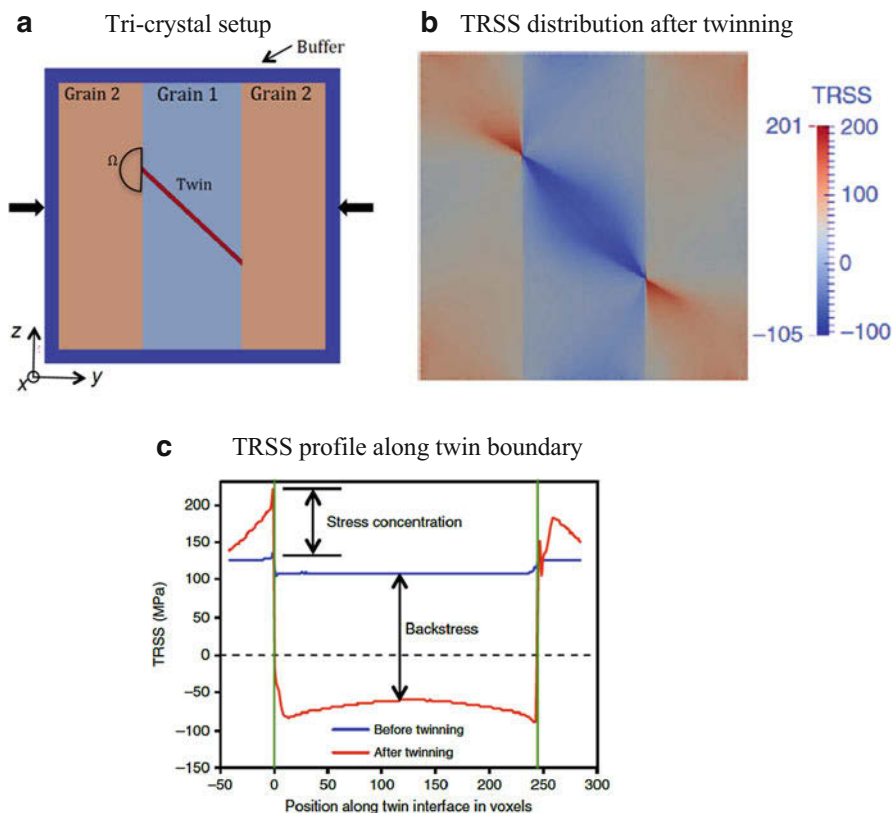


**Fig. 12** CPFFT calculated twin stress distributions generated by a twin lamella in Zr (Kumar et al. 2016a). (**a**) Tri-crystal setup consisting of a central grain "grain-1" containing a twin and two neighboring grains, with the same orientation on each side. Orientations of grain-1 and grain-2 are (0°, 0°, 0°) and (0°,30°,0°), respectively. (**b**) The twin-plane resolved shear stress (TRSS) distribution after twinning and (**c**) TRSS profile along twin boundary before and after twinning

Using CPFFT model, the stress and elastic and plastic strain tensorial fields are calculated, and for analysis, the TRSS field is presented in Fig. 12b after twinning. In this example, the neighboring grain orientation is (0°, 30°,0°) and the material is Zr. The stress inside the twin domain and along the lateral sides of the twins is negative, and in the neighboring grains, it is positive. Note that before the twin was present, the stress state was nearly uniform in the crystal, but when the twin is present, the stress state remains homogeneous in the twin domain but has become nonuniform, particularly in the matrix region immediately bordering twin. For more details, the TRSS distribution along the twin after twinning is given in Fig. 12c. The TRSS along the lateral interface of the twin band is negative, signifying that the negative TRSS along the twin boundary acts in the anti-twinning direction and thus serves as a resistance to further twinning. More quantitatively, a *twin backstress* field can be defined as the difference in the TRSS before and after twinning. The backstress is highest in value at each end, where the twin shear and reaction from the neighboring grain is the greatest and decays toward the middle of the twin. The backstress is the least at the middle of the twin. To migrate the boundaries, the applied load would need to be increased further such that the local TRSS along the twin becomes positive and exceeds a threshold value associated with boundary migration.

The amount of twin backstress observed here depends on the ability of the neighboring grain to accommodate the shear imposed by the twin. These grains deform elastically and plastically, and in the latter case, the plastic accommodation generally involves slip on multiple slip systems. For an HCP material, the activation barriers for slip depend on the mode of slip and typically include basal <a> slip, prismatic <a> slip, and pyramidal <c + a> slip. In the case of Mg, the easiest basal <a> slip (∼3 MPa) only provides two independent slip systems. Thus, basal slip alone would be insufficient to accommodate a general stress state, and in general the next easiest prismatic <a> slip and/or hardest pyramidal <c + a> slip would also have to be activated at the twin/grain boundary junction. A neighboring grain well oriented to accommodate the shearing action of the twin mostly with its easy slip system would be considered a "plastically soft" neighbor. Less backstress would result in the twin domain, particularly along the twin boundary. In contrast, a neighboring grain that is oriented poorly for easy slip must activate relatively large amounts of the harder slip modes in order to accommodate the shearing of the twin. It would constitute a "plastically hard" neighbor. A higher backstress in the twin would result and more applied load would be required to grow the twin. The grain neighbor orientation effect could be one explanation for why in a polycrystal some grains twin and others do not, despite being of similar size and orientation (Beyerlein et al. 2010; Capolungo et al. 2009; Kumar et al. 2018).

## 5.2    Grain Neighborhood Effects on Stresses to Expand the Twin

In this section, we study the effect of neighboring grain orientation on the twin backstress and twin expansion. The EVP-FFT-based twinning simulations are performed
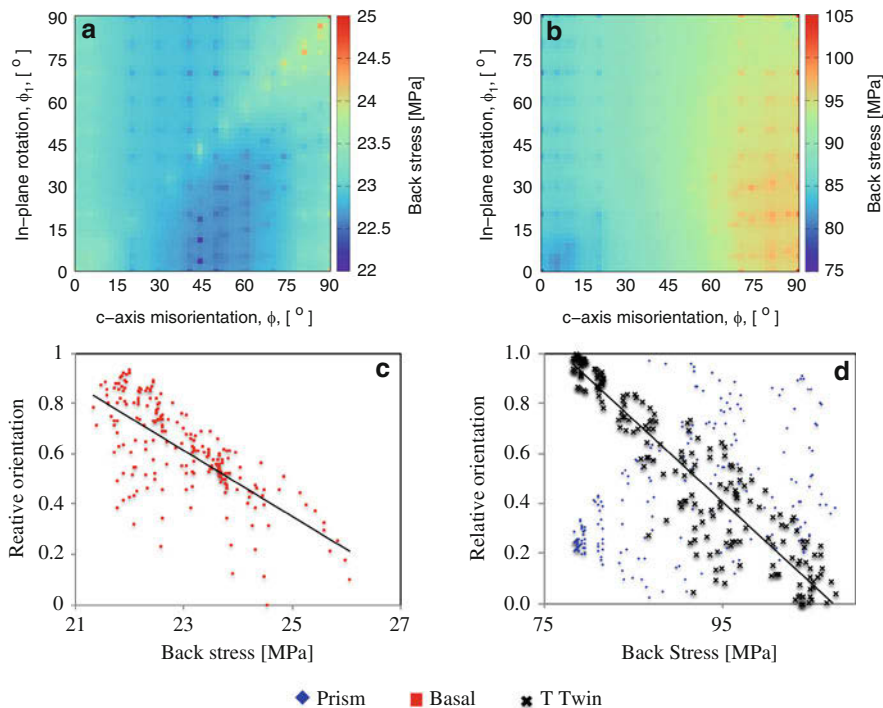
**Fig. 13** Effect of neighboring grain orientation on twin backstress (Kumar et al. 2017b). Distribution of twin backstress in the neighboring grain orientation space in (**a**) Mg and (**b**) Zr. The correlation between the twin backstress and the relative orientation of primary slip system in (**c**) Mg and (**d**) Zr. In the calculations, easy basal slip systems in Mg and easy prismatic slip and tensile twin systems in Zr are shown, and the less active systems are not shown

for 221 different grain orientations that represent the entire orientation space of neighboring grains for a fixed parent grain orientation and twin type and variant. The neighboring grain orientation space is represented in the $\phi$ versus $\phi_1$ plot, the two angles denoting the in-plane rotation and c-axis misorientation, respectively (Kumar et al. 2017b). Maps for Mg and Zr are given in Fig. 13a, b. The landscapes are very different for these two materials. The magnitudes and anisotropy in the backstress are much higher for Zr than Mg. In the case of Mg, the backstress is particularly low ($\sim$22 MPa) for the neighboring grain when the c-axis misorientation ranges from $0°$ to $\sim$25$°$ and from $\sim$65$°$ to 90$°$, but slightly higher ($\sim$25 MPa) for the c-axis misorientation range from $\sim$25$°$ to $\sim$65$°$. In the case of Zr, the backstress is lower ($\sim$75 MPa) for the c-axis misorientation range from $0°$ to $\sim$45$°$ and substantially higher ($\sim$105 MPa) for the c-axis misorientation range from $\sim$45$°$ to 90$°$.

Whether a grain neighbor is plastically hard and non-accommodating, leading to a high backstress, or vice versa, plastically soft, leading to a low one depends on the crystal's ability to activate its easiest slip mode. For Zr, the easier modes

are prismatic slip and tensile twinning, and these require higher CRSS values than, say, Mg, for which the easiest one is basal slip and the CRSS value to activate it is comparatively low. The backstress can be correlated to grain neighbor orientation through its ability to activate its easiest slip or twin mode. A simple metric that quantifies alignment between the twin in the parent and a given slip mode in the neighbor is:

$$m_{rel} = \max\left((\mathbf{b_T}.\mathbf{b_s})\,(\mathbf{n_T}.\mathbf{n_s})\right) \tag{16}$$

where $\mathbf{b_T}$ and $\mathbf{n_T}$ are the Burgers vector and plane normal unit vector of the twin and $\mathbf{b_s}$ and $\mathbf{n_s}$ are the Burgers vector and plane normal unit vector of different slip systems of neighboring grain.

This relative neighboring grain orientation factor $m_{rel}$ ranges from 0 to 1 and can be set as the maximum value calculated among the systems belonging to the slip family. A high $m_{rel}$ means that the particular deformation mode is well aligned with the twin and the likely one accommodating the twin shear. Similar measures have been defined to quantify crystallographic alignment across boundaries but for different purposes, such as slip-slip transmission (Clark et al. 1992), slip-twin transmission (Wang et al. 2010b), and twin-twin transmissions (Kumar et al. 2016a, 2017a, c) across grain boundaries.

The relationships between $m_{rel}$ for the different deformation modes and the calculated twin backstress from the CPFFT are obtained for both Mg and Zr. For pure Mg, the relationship is studied between the twin in the parent and the predominant basal slip mode in the neighbor. For Zr, the relationship is examined for both prismatic slip and the tensile twin modes. It was found that the other deformation modes are not strongly correlated with the backstress. Figure 13c shows that for Mg, the twin backstress $\tau_B$ increases as $m_{rel}$ for basal slip decreases. In the case of Zr, in Fig. 13d, the tensile twin mode exhibits the stronger correlation with $\tau_B$ than prismatic slip. In particular, the correlation between $\tau_B$ and relative orientation of neighboring grain tensile twin $m_{TTwin}$ is nearly linear following (see the line in Fig. 13d):

$$\tau_B = -26.63\,m_{TTwin} + 105.26 \tag{17}$$

This relationship although simple can be used in mean-field polycrystal models to indirectly account for neighborhood effects.

## 5.3 Incorporation of Neighborhood Effects in Homogenization Models

To simulate the role of random neighboring grain orientations on the growth of twin lamellae, the relationship expressed in Eq. (17) is incorporated into a larger scale, mean-field visco-plastic self-consistent (VPSC) model. In this hybrid VPSC model
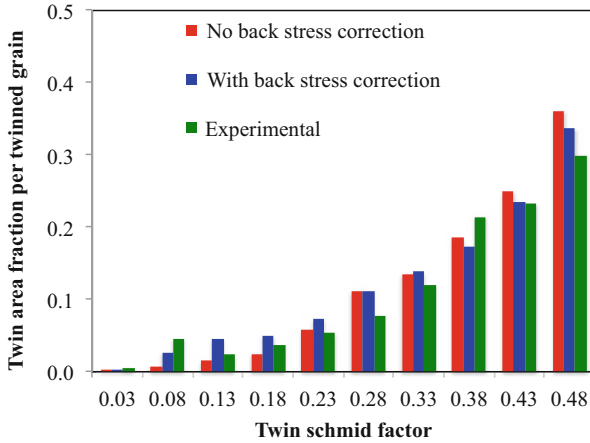
**Fig. 14** Effect of the twin backstress effect when incorporated into the VPSC model (Kumar et al. 2017b). Calculated twin area distribution from the deformation of pure Zr. Including the backstress from neighboring grains reduces twinning in the high Schmid factor region and increases it in the low Schmid factor region, aligning it better with the data than the mode without the effect of the backstress

(Kumar et al. 2017b), the twin shear rate is related to the TRSS by a power-law flow rule, which introduces a resistance to twinning:

$$\dot{\gamma}^v = \dot{\gamma}_o \left( \frac{\tau^v}{\tau_c^{\text{twin}} + \tau_B} \right)^n \tag{18}$$

where $n = 20$, $\dot{\gamma}_o$ is the reference shear rate, $\tau^v$ is the TRSS of twin variant $v$ calculated in VPSC without a neighbor effect, and $\tau_B$ is the contribution of the neighboring constraint to the resistance, the backstress we analyzed earlier with the CPFFT model.

This hybrid model is applied to polycrystalline Zr. In the simulation, for every grain, a neighboring grain orientation is randomly selected from the initial texture. Provided the grain forms a twin, its pre-assigned neighbor orientation is used to calculate the backstress on the twin using Eq. (17). Studies of the twin thickness distribution can be used to observe the effects of the neighbor on the backstress. Figure 14 compares the distribution of twin area fraction per twinned grain as a function of twin Schmid factor with the experimental measurement (see (Capolungo et al. 2009)). The model including grain neighborhood effects on twin growth reduces twinning in the high Schmid factor region yet increases it in the low Schmid factor region, providing overall better alignment with the data than the model without the effect of backstress. We find in the model that including the backstress neighbor effect lowers the growth rate *for all twins*, which is to be expected. However, in order to form the same twin fraction to accommodate the

applied strain as in the model without the backstress, the lower Schmid factor twins grew to a larger volume, an interesting consequence that again, places the hybrid model calculations in better alignment with the volume fraction of these low-rank twins.

# 6      Conclusions and Outlook

Decades of reports from experimental analyses of deformed HCP materials have undeniably demonstrated that deformation twinning is a highly heterogeneous deformation mechanism, exhibiting significant temporal and spatial variability across the crystalline microstructures. Yet to date the fundamental understanding of the mechanistic origins of the statistical and stochastic nature of these types of twins is still in development. Considering both the growing interest in HCP materials for structural applications and the profound influence of twinning on structural response, adopting a stochastic perspective in the studies of deformation twinning during mechanical deformation is sensible. In attempt to evaluate progress toward this end, this chapter reviews recent experimental analyses and modeling efforts to describe some statistical aspects of deformation twins, correlate the more statistically variable features of twins with parent microstructure, and propose the mechanisms that explain the observed microstructure/twin relationships. The contribution of grain neighborhoods to the statistical variability, as quantified by combining experimental and modeling methods, is highlighted. In addition, recent developments of stochastic twin nucleation and growth models are presented, with the aim to determine how these stochastic aspects of twinning impact mechanical behavior. The general finding is that substantial variation in twin formation, variant selection, and size significantly impacts mechanical response, from yield and hardening to ultimate strength. These strongly suggest that future pursuits for stochastic approaches to understanding the mechanical response of HCP materials that deform via twinning are worthwhile.

There are many other very important aspects of twinning that still require study from a stochastic perspective that were not covered in this chapter. Commonly seen in HCP materials, which twin profusely, are intragranular three-dimensional networks of twins. Understanding on how these twin-twin junctions form and the variation in the types of junctions that manifest would benefit from approaches that adopt statistical descriptions and stochastic models incorporated in simulation. It should also be noted that the mechanical properties reviewed in this work pertained to responses obtained in simple loading states. Twinning microstructures are, however, sensitive to deformation temperatures and imposed strain rates and deformation histories. For instance, cyclic loading can induce phenomenon such as detwinning, or changes in strain path can cause secondary twinning. There is still much opportunity for investigating the stochastic aspects of these frequently occurring twinning events.

# References

Abdolvand H, Daymond M (2013a) Multi-scale modeling and experimental study of twin inception and propagation in hexagonal close-packed materials using a crystal plasticity finite element approach-part I: average behavior. J Mech Phys Solids 61:783–802. https://doi.org/10.1016/j.jmps.2012.10.013

Abdolvand H, Daymond M (2013b) Multi-scale modeling and experimental study of twin inception and propagation in hexagonal close-packed materials using a crystal plasticity finite element approach; part II: local behavior. J Mech Phys Solids 61:803–818. https://doi.org/10.1016/j.jmps.2012.10.017

Abdolvand H, Wilkinson AJ (2016) Assessment of residual stress fields at deformation twin tips and the surrounding environments. Acta Mater 105:219–231. https://doi.org/10.1016/j.actamat.2015.11.036

Abdolvand H, Daymond M, Mareau C (2011) Incorporation of twinning into a crystal plasticity finite element model: evolution of lattice strains and texture in Zircaloy-2. Int J Plast 27:1721–1738. https://doi.org/10.1016/j.ijplas.2011.04.005

Abdolvand H et al (2015a) On the deformation twinning of Mg AZ31B: a three-dimensional synchrotron X-ray diffraction experiment and crystal plasticity finite element model. Int J Plast 70:77–97. https://doi.org/10.1016/j.ijplas.2015.03.001

Abdolvand H, Majkut M, Oddershede J, Wright J, Daymond M (2015b) Study of 3-D stress development in parent and twin pairs of a hexagonal close-packed polycrystal: part I – in-situ three-dimensional synchrotron X-ray diffraction measurement. Acta Mater 93:246–255. https://doi.org/10.1016/j.actamat.2015.04.020

Abdolvand H, Wright J, Wilkinson A (2018) Strong grain neighbour effects in polycrystals. Nat Commun 9:171. https://doi.org/10.1038/s41467-017-02213-9

Agnew SR, Brown DW, Vogel SC, Holden TM (2002) In-situ measurement of internal strain evolution during deformation dominated by mechanical twinning. In: Ecrs 6: proceedings of the 6th European conference on residual stresses, vol 404-7, pp 747–752. https://doi.org/10.4028/www.scientific.net/MSF.404-407.747

Akhtar A, Teghtsoo A (1971) Plastic deformation of Zirconium single crystals. Acta Metall Mater 19:655–663. https://doi.org/10.1016/0001-6160(71)90019-8

Ardeljan M, McCabe RJ, Beyerlein IJ, Knezevic M (2015) Explicit incorporation of deformation twins into crystal plasticity finite element models. Comput Methods Appl Mech 295:396–413. https://doi.org/10.1016/j.cma.2015.07.003

Ardeljan M, Beyerlein I, McWilliams B, Knezevic M (2016) Strain rate and temperature sensitive multi-level crystal plasticity model for large plastic deformation behavior: application to AZ31 magnesium alloy. Int J Plast 83:90–109. https://doi.org/10.1016/j.ijplas.2016.04.005

Asaro RJ (1983) Micromechanics of crystals and polycrystals. Adv Appl Mech 23:1–115. https://doi.org/10.1016/S0065-2156(08)70242-4

Asaro RJ, Lubarda VA (2006) Mechanics of solids and materials. Cambridge University Press, Cambridge/New York

Barnett MR (2007a) Twinning and the ductility of magnesium alloys part I: "Tension" twins. Mater Sci Eng A-Struct 464:1–7. https://doi.org/10.1016/j.msea.2006.12.037

Barnett MR (2007b) Twinning and the ductility of magnesium alloys part II. "Contraction" twins. Mater Sci Eng A-Struct 464:8–16. https://doi.org/10.1016/j.msea.2007.02.109

Barnett MR, Keshavarz Z, Beer AG, Atwell D (2004) Influence of grain size on the compressive deformation of wrought Mg-3Al-1Zn. Acta Mater 52:5093–5103. https://doi.org/10.1016/j.actamat.2004.07.015

Barnett MR, Nave MD, Ghaderi A (2012) Yield point elongation due to twinning in a magnesium alloy. Acta Mater 60:1433–1443. https://doi.org/10.1016/j.actamat.2011.11.022

Barrett CD, El Kadiri H (2014) The roles of grain boundary dislocations and disclinations in the nucleation of {10(1)over-bar12} twinning. Acta Mater 63:1–15. https://doi.org/10.1016/j.actamat.2013.09.012

Bay B, Hansen N, Hughes DA, Kuhlmannwilsdorf D (1992) Overview No-96 – evolution of F.C.C. deformation structures in Polyslip. Acta Metall Mater 40:205–219. https://doi.org/10.1016/0956-7151(92)90296-Q

Beyerlein IJ, Tome CN (2008) A dislocation-based constitutive law for pure Zr including temperature effects. Int J Plast 24:867–895. https://doi.org/10.1016/j.ijplas.2007.07.017

Beyerlein IJ, Tome CN (2010) A probabilistic twin nucleation model for HCP polycrystalline metals. Proc R Soc A-Math Phys 466:2517–2544. https://doi.org/10.1098/rspa.2009.0661

Beyerlein IJ, Capolungo L, Marshall PE, McCabe RJ, Tome CN (2010) Statistical analyses of deformation twinning in magnesium. Philos Mag 90:2161. https://doi.org/10.1080/14786435.2010.517966. 90:4073-4074. Pii 926495039

Beyerlein IJ, McCabe RJ, Tome CN (2011) Effect of microstructure on the nucleation of deformation twins in polycrystalline high-purity magnesium: a multi-scale modeling study. J Mech Phys Solids 59:988–1003. https://doi.org/10.1016/j.jmps.2011.02.007

Beyerlein IJ, Wang J, Barnett MR, Tome CN (2012) Double twinning mechanisms in magnesium alloys via dissociation of lattice dislocations. Proc R Soc A-Math Phys 468:1496–1520. https://doi.org/10.1098/rspa.2011.0731

Beyerlein IJ, Wang J, Kang K, Zheng SJ, Mara NA (2013) Twinnability of bimetal interfaces in nanostructured composites. Mater Res Lett 1:89–95. https://doi.org/10.1080/21663831.2013.782074

Beyerlein IJ, Zhang XH, Misra A (2014) Growth twins and deformation twins in metals. Annu Rev Mater Res 44:329–363. https://doi.org/10.1146/annurev-matsci-070813-113304

Bieler T, Wang L, Beaudoin A, Kenesei P, Lienert U (2014) In situ characterization of twin nucleation in pure Ti using 3D-XRD. Metall Mater Trans A 45A:109–122. https://doi.org/10.1007/s11661-013-2082-3

Brenner R, Lebensohn RA, Castelnau O (2009) Elastic anisotropy and yield surface estimates of polycrystals. Int J Solids Struct 46:3018–3026. https://doi.org/10.1016/j.ijsolstr.2009.04.001

Bronkhorst CA, Kalidindi SR, Anand L (1992) Polycrystalline plasticity and the evolution of crystallographic texture in FCC metals. Philos Trans R Soc A 341:443–477. https://doi.org/10.1098/rsta.1992.0111

Canova GR, Fressengeas C, Molinari A, Kocks UF (1988) Effect of rate sensitivity on slip system activity and lattice rotation. Acta Metall Mater 36:1961–1970. https://doi.org/10.1016/0001-6160(88)90298-2

Capolungo L, Beyerlein IJ (2008) Nucleation and stability of twins in hcp metals. Phys Rev B 78:024117. https://doi.org/10.1103/PhysRevB.78.024117. ARTN 024117

Capolungo L, Marshall PE, McCabe RJ, Beyerlein IJ, Tome CN (2009) Nucleation and growth of twins in Zr: a statistical study. Acta Mater 57:6047–6056. https://doi.org/10.1016/j.actamat.2009.08.030

Cheng JH, Ghosh S (2015) A crystal plasticity FE model for deformation with twin nucleation in magnesium alloys. Int J Plast 67:148–170. https://doi.org/10.1016/j.ijplas.2014.10.005

Cheng JH, Ghosh S (2017a) Crystal plasticity finite element modeling of discrete twin evolution in polycrystalline magnesium. J Mech Phys Solids 99:512–538. https://doi.org/10.1016/j.jmps.2016.12.008

Cheng JH, Ghosh S (2017b) Simulating discrete twin evolution in magnesium using a novel crystal plasticity finite element model. Miner Met Mater Ser:167–174. https://doi.org/10.1007/978-3-319-52392-7_26

Clark WAT, Wagoner RH, Shen ZY, Lee TC, Robertson IM, Birnbaum HK (1992) On the criteria for slip transmission across interfaces in polycrystals. Scr Metall Mater 26:203–206. https://doi.org/10.1016/0956-716x(92)90173-C

Cohen JB, Weertman J (1963) A dislocation model for twinning in F.C.C. Metals. Acta Metall Mater 11:996. https://doi.org/10.1016/0001-6160(63)90074-9

Delannay L, Jacques PJ, Kalidindi SR (2006) Finite element modeling of crystal plasticity with grains shaped as truncated octahedrons. Int J Plast 22:1879–1898. https://doi.org/10.1016/j.ijplas.2006.01.008

Ecob N, Ralph B (1983) The effect of grain-size on deformation twinning in a textured zinc alloy. J Mater Sci 18:2419–2429. https://doi.org/10.1007/Bf00541848

Eisenlohr P, Diehl M, Lebensohn RA, Roters F (2013) A spectral method solution to crystal elasto-viscoplasticity at finite strains. Int J Plast 46:37–53. https://doi.org/10.1016/j.ijplas.2012.09.012

Fisher ES, Renken CJ (1964) Single-crystal elastic moduli + hcp to bcc transformation in Ti,Zr, and Hf. Phys Rev 135:A482–A494. https://doi.org/10.1103/PhysRev.135.A482

Ghaderi A, Barnett M (2011) Sensitivity of deformation twinning to grain size in titanium and magnesium. Acta Mater 59:7824–7839. https://doi.org/10.1016/j.actamat.2011.09.018

Gutierrez-Urrutia I, Raabe D (2012) Grain size effect on strain hardening in twinning-induced plasticity steels. Scr Mater 66:992–996. https://doi.org/10.1016/j.scriptamat.2012.01.037

Hazeli K, Cuadra J, Vanniamparambil PA, Kontsos A (2013) In situ identification of twin-related bands near yielding in a magnesium alloy. Scr Mater 68:83–86. https://doi.org/10.1016/j.scriptamat.2012.09.009

Hearmon RFS (1946) The elastic constants of anisotropic materials. Rev Mod Phys 18:409–440. https://doi.org/10.1103/RevModPhys.18.409

Hosford WF (1993) The mechanics of crystals and textured polycrystals. The Oxford engineering science series, vol 32. Oxford University Press, New York

Hughes DA, Hansen N (1997) High angle boundaries formed by grain subdivision mechanisms. Acta Mater 45:3871–3886. https://doi.org/10.1016/S1359-6454(97)00027-X

Idiart MI, Moulinec H, Castaneda PP, Suquet P (2006) Macroscopic behavior and field fluctuations in viscoplastic composites: second-order estimates versus full-field simulations. J Mech Phys Solids 54:1029–1063. https://doi.org/10.1016/j.jmps.2005.11.004

Jagannadham K (1976) Lattice dislocation near a surface. J Appl Phys 47:4401–4403. https://doi.org/10.1063/1.322445

Jain A, Duygulu O, Brown DW, Tome CN, Agnew SR (2008) Grain size effects on the tensile properties and deformation mechanisms of a magnesium alloy, AZ31B, sheet. Mat Sci Eng A-Struct 486:545–555. https://doi.org/10.1016/j.msea.2007.09.069

Juan PA, Pradalier C, Berbenni S, McCabe RJ, Tome CN, Capolungo L (2015) A statistical analysis of the influence of microstructure and twin-twin junctions on twin nucleation and twin growth in Zr. Acta Mater 95:399–410. https://doi.org/10.1016/j.actamat.2015.05.022

Kacher J, Minor AM (2014) Twin boundary interactions with grain boundaries investigated in pure rhenium. Acta Mater 81:1–8. https://doi.org/10.1016/j.actamat.2014.08.013

Kalidindi SR (1998) Incorporation of deformation twinning in crystal plasticity models. J Mech Phys Solids 46:267–290. https://doi.org/10.1016/S0022-5096(97)00051-3

Kang S, Jung JG, Kang M, Woo W, Lee YK (2016) The effects of grain size on yielding, strain hardening, and mechanical twinning in Fe-18Mn-0.6C-1.5Al twinning-induced plasticity steel. Mat Sci Eng A-Struct 652:212–220. https://doi.org/10.1016/j.msea.2015.11.096

Kanjarla AK, Beyerlein IJ, Lebensohn RA, Tome CN (2012a) On the role of local grain interactions on twin nucleation and texture evolution in hexagonal materials. Mater Sci Forum 702-703:265–268. https://doi.org/10.4028/www.scientific.net/MSF.702-703.265

Kanjarla AK, Lebensohn RA, Balogh L, Tome CN (2012b) Study of internal lattice strain distributions in stainless steel using a full-field elasto-viscoplastic formulation based on fast Fourier transforms. Acta Mater 60:3094–3106. https://doi.org/10.1016/j.actamat.2012.02.014

Kaschner GC, Tome CN, Beyerlein IJ, Vogel SC, Brown DW, McCabe RJ (2006) Role of twinning in the hardening response of zirconium during temperature reloads. Acta Mater 54:2887–2896. https://doi.org/10.1016/j.actamat.2006.02.036

Khosravani A, Fullwood DT, Adams BL, Rampton TM, Miles MP, Mishra RK (2015) Nucleation and propagation of {1 0(1)over-bar2} twins in AZ31 magnesium alloy. Acta Mater 100:202–214. https://doi.org/10.1016/j.actamat.2015.08.024

Knezevic M, Zecevic M, Beyerlein I, Bingert J, McCabe R (2015) Strain rate and temperature effects on the selection of primary and secondary slip and twinning systems in HCP Zr. Acta Mater 88:55–73. https://doi.org/10.1016/j.actamat.2015.01.037

Knezevic M, Zecevic M, Beyerlein I, Lebensohn R (2016) A numerical procedure enabling accurate descriptions of strain rate-sensitive flow of polycrystals within crystal visco-plasticity theory. Comput Method Appl Mech 308:468–482. https://doi.org/10.1016/j.cma.2016.05.025

Kocks UF (1970) Relation between polycrystal deformation and single-crystal deformation. Metall Trans 1:1121–1143. https://doi.org/10.1007/Bf02900224

Kocks UF, Tomé CN, Wenk HR (2000) Texture and anisotropy – preferred orientations in polycrystals and their effect on materials properties. Cambridge University Press, Cambridge

Kuhlmann-Wilsdorf D (1999) Overview no. 131 – "Regular" deformation bands (DBs) and the LEDS hypothesis. Acta Mater 47:1697–1712. https://doi.org/10.1016/S1359-6454(98)00413-3

Kuhlman-wilsdorf D, Hansen N (1991) Geometrically necessary, incidental and subgrain boundaries. Scr Metall Mater 25:1557–1562. https://doi.org/10.1016/0956-716x(91)90451-6

Kumar MA, Mahesh S (2012) Banding in single crystals during plastic deformation. Int J Plast 36:15–33. https://doi.org/10.1016/j.ijplas.2012.03.008

Kumar MA, Kanjarla AK, Niezgoda SR, Lebensohn RA, Tome CN (2015) Numerical study of the stress state of a deformation twin in magnesium. Acta Mater 84:349–358. https://doi.org/10.1016/j.actamat.2014.10.048

Kumar MA, Beyerlein IJ, McCabe RJ, Tome CN (2016a) Grain neighbour effects on twin transmission in hexagonal close-packed materials. Nat Commun 7:13826 https://doi.org/10.1038/ncomms13826

Kumar MA, Beyerlein IJ, Tome CN (2016b) Effect of local stress fields on twin characteristics in HCP metals. Acta Mater 116:143–154. https://doi.org/10.1016/j.actamat.2016.06.042

Kumar MA, Beyerlein IJ, Tome CN (2016c) Grain size constraints on twin expansion in hexagonal close packed crystals. J Appl Phys 120. https://doi.org/10.1063/1.4965719. Artn 155105

Kumar MA, Beyerlein IJ, Lebensohn RA, Tome CN (2017a) Modeling the effect of alloying elements in magnesium on deformation twin characteristics. Miner Met Mater Ser:159–165. https://doi.org/10.1007/978-3-319-52392-7_25

Kumar MA, Beyerlein IJ, Lebensohn RA, Tome CN (2017b) Modeling the effect of neighboring grains on twin growth in HCP polycrystals. Model Simul Mater Sci 25. https://doi.org/10.1088/1361-651X/aa7bbb. ARTN 064007

Kumar MA, Beyerlein IJ, Lebensohn RA, Tome CN (2017c) Role of alloying elements on twin growth and twin transmission in magnesium alloys. Mat Sci Eng A-Struct 706:295–303. https://doi.org/10.1016/j.msea.2017.08.084

Kumar MA, Beyerlein IJ, Tome CN (2017d) A measure of plastic anisotropy for hexagonal close packed metals: application to alloying effects on the formability of Mg. J Alloys Compd 695:1488–1497. https://doi.org/10.1016/j.jallcom.2016.10.287

Kumar MA, Wroński M, McCabe RJ, Capolungo L, Wierzbanowski K, Tomé CN (2018) Role of microstructure on twin nucleation and growth in HCP titanium: a statistical study. Acta Mater 148:123–132

Lebensohn RA (2001) N-site modeling of a 3D viscoplastic polycrystal using fast Fourier transform. Acta Mater 49:2723–2737

Lebensohn RA, Tome CN (1993) A self-consistent anisotropic approach for the simulation of plastic-deformation and texture development of polycrystals – application to zirconium alloys. Acta Metall Mater 41:2611–2624. https://doi.org/10.1016/0956-7151(93)90130-K

Lebensohn RA, Brenner R, Castelnau O, Rollett AD (2008) Orientation image-based microme-
    chanical modelling of subgrain texture evolution in polycrystalline copper. Acta Mater 56:3914–
    3926. https://doi.org/10.1016/j.actamat.2008.04.016
Lebensohn RA, Castañeda PP, Brenner R, Castelnau O (2011a) Full-field vs. homogenization
    methods to predict microstructure–property relations for polycrystalline materials. In: Ghosh S,
    Dimiduk D (eds) Computational methods for microstructure-property relationships. Springer,
    Boston
Lebensohn RA, Idiart MI, Castaneda PP, Vincent PG (2011b) Dilatational viscoplasticity of
    polycrystalline solids with intergranular cavities. Philos Mag 91:3038–3067. https://doi.org/10.
    1080/14786435.2011.561811
Lebensohn RA, Kanjarla AK, Eisenlohr P (2012) An elasto-viscoplastic formulation based on fast
    Fourier transforms for the prediction of micromechanical fields in polycrystalline materials. Int
    J Plast 32–33:59–69. https://doi.org/10.1016/j.ijplas.2011.12.005
Lentz M, Behringer A, Fahrenson C, Beyerlein IJ, Reimers W (2014) Grain size effects on primary,
    secondary, and tertiary twin development in Mg-4 wt pct Li (-1 wt pct Al) alloys. Metall Mater
    Trans A 45a:4737–4741. https://doi.org/10.1007/s11661-014-2491-y
Lentz M, Klaus M, Beyerlein IJ, Zecevic M, Reimers W, Knezevic M (2015) In situ X-ray
    diffraction and crystal plasticity modeling of the deformation behavior of extruded Mg-Li-(Al)
    alloys: an uncommon tension-compression asymmetry. Acta Mater 86:254–268. https://doi.org/
    10.1016/j.actamat.2014.12.003
Lentz M, Risse M, Schaefer N, Reimers W, Beyerlein IJ (2016) Strength and ductility with
    {10(1)over-bar1} – {10(1)over-bar2} double twinning in a magnesium alloy. Nat Commun 7.
    https://doi.org/10.1038/ncomms11068. ARTN 11068
Liu B, Raabe D, Roters F, Eisenlohr P, Lebensohn RA (2010) Comparison of finite element and
    fast Fourier transform crystal plasticity solvers for texture prediction. Model Simul Mater Sci
    18. https://doi.org/10.1088/0965-0393/18/8/085005. Artn 085005
Liu X, Nuhfer NT, Warren AP, Coffey KR, Rohrer GS, Barmak K (2015) Grain size dependence of
    the twin length fraction in nanocrystalline Cu thin films via transmission electron microscopy
    based orientation mapping. J Mater Res 30:528–537. https://doi.org/10.1557/jmr.2014.393
Mahajan S, Chin GY (1973) Formation of deformation twins in Fcc crystals. Acta Metall Mater
    21:1353–1363. https://doi.org/10.1016/0001-6160(73)90085-0
Masson R, Bornert M, Suquet P, Zaoui A (2000) An affine formulation for the prediction of the
    effective properties of nonlinear composites and polycrystals. J Mech Phys Solids 48:1203–
    1227. https://doi.org/10.1016/S0022-5096(99)00071-X
McCabe RJ, Proust G, Cerreta EK, Misra A (2009) Quantitative analysis of deformation twinning
    in zirconium. Int J Plast 25:454–472. https://doi.org/10.1016/j.ijplas.2008.03.010
Mendelson S (1972) Dislocation dissociations and dislocation mobility in diamond lattice crystals.
    J Appl Phys 43:2102. https://doi.org/10.1063/1.1661460
Michel J, Moulinec H, Suquet P (2000) A computational method based on augmented Lagrangians
    and fast Fourier transforms for composites with high contrast. Comput Model Eng Sci 1:79–88
Michel J, Moulinec H, Suquet P (2001) A computational scheme for linear and non-linear
    composites with arbitrary phase contrast. Int J Numer Methods Eng 52:139–158. https://doi.
    org/10.1002/nme.275
Mika D, Dawson P (1999) Polycrystal plasticity modeling of intracrystalline boundary textures.
    Acta Mater 47:1355–1369. https://doi.org/10.1016/S1359-6454(98)00386-3
Morrow BM, Cerreta EK, McCabe RJ, Tome CN (2014a) Toward understanding twin-twin interac-
    tions in hcp metals: utilizing multiscale techniques to characterize deformation mechanisms in
    magnesium. Mater Sci Eng A-Struct 613:365–371. https://doi.org/10.1016/j.msea.2014.06.062
Morrow BM, Mccabe RJ, Cerreta EK, Tome CN (2014b) In-situ TEM observation of twinning and
    detwinning during cyclic loading in Mg. Metall Mater Trans A 45a:36–40. https://doi.org/10.
    1007/s11661-013-1765-0
Morrow BM, McCabe RJ, Cerreta EK, Tome CN (2014c) Observations of the atomic structure of
    tensile and compressive twin boundaries and twin-twin interactions in zirconium. Metall Mater
    Trans A 45a:5891–5897. https://doi.org/10.1007/s11661-014-2481-0

Moulinec H, Suquet P (1994) A fast numerical method for computing the linear and nonlinear mechanical properties of composites. C R Acad Sci Ser II 318: 1417–1423

Moulinec H, Suquet P (1998) A numerical method for computing the overall response of nonlinear composites with complex microstructure. Comput Method Appl Mech 157:69–94. https://doi.org/10.1016/S0045-7825(97)00218-1

Niezgoda SR, Beyerlein IJ, Kanjarla AK, Tome CN (2013) Introducing grain boundary influenced stochastic effects into constitutive models. Jom-Us 65:419–430. https://doi.org/10.1007/s11837-012-0550-7

Niezgoda SR, Kanjarla AK, Beyerlein IJ, Tome CN (2014) Stochastic modeling of twin nucleation in polycrystals: an application in hexagonal close-packed metals. Int J Plast 56:119–138. https://doi.org/10.1016/j.ijplas.2013.11.005

Partridge PG (1967) The crystallography and deformation modes of hexagonal close-packed metals. Metall Rev 12:169–194

Peirce D, Asaro RJ, Needleman A (1982) An analysis of nonuniform and localized deformation in ductile single-crystals. Acta Metall Mater 30:1087–1119. https://doi.org/10.1016/0001-6160(82)90005-0

Priestner R, Leslie WC (1965) Nucleation of deformation twins at slip plane intersections in Bcc metals. Philos Mag 11:895–916. https://doi.org/10.1080/14786436508223953

Proust G, Tome CN, Kaschner GC (2007) Modeling texture, twinning and hardening evolution during deformation of hexagonal materials. Acta Mater 55:2137–2148. https://doi.org/10.1016/j.actamat.2006.11.017

Proust G, Tome CN, Jain A, Agnew SR (2009) Modeling the effect of twinning and detwinning during strain-path changes of magnesium alloy AZ31. Int J Plast 25:861–880. https://doi.org/10.1016/j.ijplas.2008.05.005

Rahman KM, Vorontsov VA, Dye D (2015) The effect of grain size on the twin initiation stress in a TWIP steel. Acta Mater 89:247–257. https://doi.org/10.1016/j.actamat.2015.02.008

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58:1152–1211. https://doi.org/10.1016/j.actamat.2009.10.058

Salem AA, Kalidindi SR, Doherty RD, Semiatin SL (2006) Strain hardening due to deformation twinning in alpha-titanium: Mechanisms. Metall Mater Trans A 37a:259–268. https://doi.org/10.1007/s11661-006-0171-2

Shanthraj P, Eisenlohr P, Diehl M, Roters F (2015) Numerically robust spectral methods for crystal plasticity simulations of heterogeneous materials. Int J Plast 66:31–45. https://doi.org/10.1016/j.ijplas.2014.02.006

Shi ZZ et al (2015a) On the selection of extension twin variants with low Schmid factors in a deformed Mg alloy. Acta Mater 83:17–28. https://doi.org/10.1016/j.actamat.2014.10.004

Shi ZZ et al (2015b) Variant selection of twins with low Schmid factors in cross grain boundary twin pairs in a magnesium alloy. Iop Conf Ser-Mat Sci 82. https://doi.org/10.1088/1757-899x/82/1/012021. Artn 012021

Shi ZZ et al (2015c) Sequential double extension twinning in a magnesium alloy: combined statistical and micromechanical analyses. Acta Mater 96:333–343. https://doi.org/10.1016/j.actamat.2015.06.029

Simkin BA, Ng BC, Crimp MA, Bieler TR (2007) Crack opening due to deformation twin shear at grain boundaries in near-gamma TiAl. Intermetallics 15:55–60. https://doi.org/10.1016/j.intermet.2006.03.005

Simmons G, Wang H (1971) Single crystal elastic constants and calculated aggregate properties: a handbook. MIT press, Cambridge, MA

Stanford N, Barnett MR (2008) Fine grained AZ31 produced by conventional thermo-mechanical processing. J Alloys Compd 466:182–188. https://doi.org/10.1016/j.jallcom.2007.11.082

Thorvaldsen A (1993) Grain-growth as a stochastic process. Acta Metall Mater 41:1347–1357. https://doi.org/10.1016/0956-7151(93)90244-M

Tome C, Canova GR, Kocks UF, Christodoulou N, Jonas JJ (1984) The relation between macroscopic and microscopic strain-hardening in Fcc polycrystals. Acta Metall Mater 32:1637–1653. https://doi.org/10.1016/0001-6160(84)90222-0

Tome C, Lebensohn R, Kocks U (1991) A model for texture development dominated by deformation twinning – application to zirconium alloys. Acta Metall Mater 39:2667–2680. https://doi.org/10.1016/0956-7151(91)90083-D

Tsai MS, Chang CP (2013) Grain size effect on deformation twinning in Mg-Al-Zn alloy. Mater Sci Technol-Lond 29:759–763. https://doi.org/10.1179/1743284713y.0000000237

Wang H, Wu PD, Wang J, Tome CN (2013a) A crystal plasticity model for hexagonal close packed (HCP) crystals including twinning and de-twinning mechanisms. Int J Plast 49:36–52. https://doi.org/10.1016/j.ijplas.2013.02.016

Wang J, Beyerlein IJ, Tome CN (2010a) An atomic and probabilistic perspective on twin nucleation in Mg. Scr Mater 63:741–746. https://doi.org/10.1016/j.scriptamat.2010.01.047

Wang J, Beyerlein IJ, Hirth JP, Tome CN (2011) Twinning dislocations on {(1)over-bar011} and {(1)over-bar013} planes in hexagonal close-packed crystals. Acta Mater 59:3990–4001. https://doi.org/10.1016/j.actamat.2011.03.024

Wang J, Beyerlein IJ, Hirth JP (2012) Nucleation of elementary {(1)over-bar 0 1 1} and {(1)over-bar 0 1 3} twinning dislocations at a twin boundary in hexagonal close-packed crystals. Model Simul Mater Sci 20:024001. https://doi.org/10.1088/0965-0393/20/2/024001

Wang J, Yadav SK, Hirth JP, Tome CN, Beyerlein IJ (2013b) Pure-shuffle nucleation of deformation twins in hexagonal-close-packed metals. Mater Res Lett 1:126–132. https://doi.org/10.1080/21663831.2013.792019

Wang J, Beyerlein IJ, Tome CN (2014) Reactions of lattice dislocations with grain boundaries in Mg: implications on the micro scale from atomic-scale calculations. Int J Plast 56:156–172. https://doi.org/10.1016/j.ijplas.2013.11.009

Wang L, Yang Y, Eisenlohr P, Bieler TR, Crimp MA, Mason DE (2010b) Twin nucleation by slip transfer across grain boundaries in commercial purity titanium. Metall Mater Trans A 41a:421–430. https://doi.org/10.1007/s11661-009-0097-6

Wang LY, Barabash R, Bieler T, Liu WJ, Eisenlohr P (2013c) Study of {1211} twinning in α-Ti by EBSD and Laue microdiffraction. Metall Mater Trans A 44:3664–3674. https://doi.org/10.1007/s11661-013-1714-y

Wongwiwat K, Murr LE (1978) Effect of shock pressure, pulse duration, and grain-size on shock-deformation twinning in molybdenum. Mater Sci Eng 35:273–285. https://doi.org/10.1016/0025-5416(78)90129-5

Wronski M, Arul Kumar M, Capolungo L, Madec R, Wierzbanowski K, Tome CN (2018) Deformation behavior of CP-titanium: experiment and crystal plasticity modeling. Mater Sci Eng A-Struct 724:289–297

Wu W, Chuang CP, Qiao DX, Ren Y, An K (2016) Investigation of deformation twinning under complex stress states in a rolled magnesium alloy. J Alloys Compd 683:619–633. https://doi.org/10.1016/j.jallcom.2016.05.144

Yang F, Yin SM, Li SX, Zhang ZF (2008) Crack initiation mechanism of extruded AZ31 magnesium alloy in the very high cycle fatigue regime. Mater Sci Eng A-Struct 491:131–136. https://doi.org/10.1016/j.msea.2008.02.003

Yin SM, Yang F, Yang XM, Wu SD, Li SX, Li GY (2008) The role of twinning-detwinning on fatigue fracture morphology of Mg-3%Al-1%Zn alloy. Mater Sci Eng A-Struct 494:397–400. https://doi.org/10.1016/j.msea.2008.04.056

Yoo MH (1981) Slip, twinning, and fracture in hexagonal close-packed metals. Metall Trans A 12:409–418. https://doi.org/10.1007/Bf02648537

Yoo MH, Lee JK (1991) Deformation twinning in Hcp metals and alloys. Philos Mag A 63:987–1000. https://doi.org/10.1080/01418619108213931

Zecevic M, Knezevic M (2017) Modeling of sheet metal forming based on implicit embedding of the elasto-plastic self-consistent formulation in shell elements: application to cup drawing of AA6022-T4. JOM 69:922–929. https://doi.org/10.1007/s11837-017-2255-4

Zhao Z, Kuchnicki S, Radovitzky R, Cultino A (2007) Influence of in-grain mesh resolution on the prediction of deformation textures in fcc polycrystals by crystal plasticity FEM. Acta Mater 55:2361–2373. https://doi.org/10.1016/j.actamat.2006.11.035

Zheng SJ, Beyerlein IJ, Wang J, Carpenter JS, Han WZ, Mara NA (2012) Deformation twinning mechanisms from bimetal interfaces as revealed by in situ straining in the TEM. Acta Mater 60:5858–5866. https://doi.org/10.1016/j.actamat.2012.07.027

# Part IX
# Soft Matter/Polymer Simulations

# Soft Matter/Polymer Simulations and Bridging Scales: Overview

# 60

Kurt Kremer

## Contents

**Abstract**

This collection of papers gives an introduction and overview of recent developments in polymer/soft matter simulations. The variety of the contributions reflects the fact that soft matter not only is based on a fast-growing body of elementary chemical structures but even more so on the way these structures are organized and interact with each other resulting in their specific structure/morphology and function.

## 1    Materials Properties and Link to Statistical Mechanics

A comprehensive polymer or – better – soft materials modeling, preferentially based on a thorough theoretical understanding of the underlying physical principles and chemical structure, requires to consider properties and processes originating from many connected length and time scales. Ideally that would mean a seamless modeling strategy from quantum chemical methods all the way to macroscopic continuum-based methodologies. Furthermore almost all experimentally available systems are the result of a specific preparation procedure, meaning that their state

K. Kremer (✉)
Max Planck Institute for Polymer Research, Mainz, Germany
e-mail: kremer@mpip-mainz.mpg.de; theoryoffice@mpip-mainz.mpg.de

is history dependent and thus not in thermodynamic equilibrium (in terms of the physical definition). Systems are "stuck" in a local free energy minimum. This requires a modified view away from the traditional structure property relationship (SPR) toward a structure-*process*-property relationship (SPPR), based on insight into processes on a molecular level. This holds even more so for all biological systems.

Polymer/soft matter physics flourished with the development of universality and scaling relationships for polymers in the last century by de Gennes, Edwards, and many others (de Gennes 1979; Doi and Edwards 1986; Rubinstein and Colby 2003; Doi 2013). That connected soft matter to the most modern statistical physics at that time. Polymers were not only materials with interesting physical properties, but they also became means to study modern physics. Scaling laws provided a generic understanding of whole classes of materials and allowed a systematic theoretical treatment. In parallel computer simulations, which focused on such generic models, began to become more widely possible. At the same time, chemistry similarly advanced and provided a huge wealth of new materials with rather different properties. Mechanical, chemical, and electronic properties and functionality can vary dramatically for different chemical species while the generic laws still hold. A rather trivial example may illustrate this. The viscosity $\eta$ of a polymer melt follows the law $\eta = f(T, \dots) N^{3.4}$, N being the polymerization index of the chains and f being a function of temperature, local chemistry, etc.. Doubling N increases the viscosity by roughly a factor of 10. This holds for all polymers, i.e., it is universally valid. However by, e.g., changing the temperature by a few degrees close to the glass transition temperature, $\eta$ also easily changes by a factor of 10 or more. Thus the prefactor f and the power law allow the adjustment of materials properties to the same degree, however by different means. It is this extreme versatility based on a combination of chemistry-specific aspects and generic power laws that makes soft matter an ubiquitously demanded class of materials (Potestio et al. 2014). Modern materials modeling has to take these two adjusting screws properly into account. Needless to say, that despite all the progress over the last years, many old and new challenges are tackled by the community.

## 2    Contributions

The collection of articles in this chapter provides an overview of recent developments in different areas of computational physics and materials science of soft matter. It covers topics from classical polymer solutions and melt materials simulations all the way to modern data-driven approaches. While such a collection cannot be complete, it should provide insight and guidance to further literature.

▶ Chapter 61, "Polymer Solutions" by B. Dünweg builds upon scaling relationships for statics and dynamics. It introduces the intuitive blob picture and the crossover scaling concept. While being especially of interest in the context of theoretical physics, polymer solutions play an important role in many areas of

technology, e.g., as viscosity modifiers. After introducing the theoretical concepts, the power of modern advanced simulation approaches is illustrated. This includes a short discussion of the difference between simulations including explicit versus implicit solvent. While less relevant for actual materials science problems, concepts and technologies developed in the context of solutions form a starting point for many applications presented in the subsequent chapters.

A special class of polymers with ample applications in science and technology ranging from chemical industry through health care and food science, not to speak of their biological relevance, are polyelectrolytes. In ▶ Chap. 62, "From the Atomistic to the Macromolecular Scale: Distinct Simulation Approaches for Polyelectrolyte Solutions," J. Smiatek and Ch. Holm cover the range of simulations from systems with atomistic details to more coarse generic models allowing to study structure formation and aggregation. Polyelectrolytes are charged macromolecules, where charges are dissociated from (some of) the beads. They can, depending on chemistry, either be anionic or cationic. In some special cases, one can have both types of charges within one macromolecule. Strong polyelectrolytes are completely ionized, while weak polyelectrolytes are only partially ionized. The degree of ionization, e.g., depends not only on pH but also on the properties of the uncharged parts of the molecule in between. Material properties equally depend on the long-range properties of the electrostatic interaction, i.e., screening due to salt, etc., as well as local chemistry-dependent aspects like poor solvent collapse or aggregation and ion dissociation.

At the other end of classical polymer materials problems are long chain polymer melts. ▶ Chapter 63, "Resolving Properties of Entangled Polymers Melts Through Atomistic Derived Coarse-Grained Models" by G. S. Grest and coworkers explain in detail how to obtain quantitative information for the dynamics in polymer melts by particle-based multiscale simulations. As an example, the most abundant commodity polymer polyethylene (PE) is chosen. This is the most simple polymer – a chain of $CH_2$ groups – which, however, displays all the characteristics and complications of conventional polymer melts. Thus it provides a perfect demonstration case as, on the one hand, local energy barriers and interactions are well known, making it possible to connect simulations to experiment without any free parameter. On the other hand, very precise experiments are available. The authors in detail explain how different levels of coarse graining affect the dynamics of the coarse-grained model in comparison to the underlying all-atom model. While the mapping of length scales is trivially given by the mapping procedure between all-atom and coarse-grained beads, this is not at all the case for dynamics. Actually mapping dynamics properties between different models is not at all trivial and still poses one of the central questions to be solved. This is a currently highly active research topic at the interface between applied mathematics and materials modeling. However, it would be beyond the scope of this chapter to cover such still rather preliminary research at this point.

So far, the chapters mentioned above described the classical approach of a bottom-up ansatz. Starting from a detailed, eventually all-atom description, systematically coarse-grained models are developed and studied. A complementary way of

incorporating specific materials properties into coarse models is given in the ▶ Chap. 64, "Top-Down Hybrid Models of Polymers" by K. Daoulas. Within these models, particles interact with their surrounding not in a particle-particle fashion but with an averaged field formed by all the other surrounding particles. This can speed up simulations dramatically; however the methodology includes several complications, which have to be addressed carefully. In this chapter, the theoretical framework is discussed in detail, and a solid basis for practical work is provided. Including specific materials parameters into these rather coarse models also requires special care. Recently these methods have found ample applications in block copolymer systems and polymer systems with locally anisotropic morphology, such as dense conjugated polymers studied for organics electronics.

When dealing with polymer electronics, all the above-discussed methodologies come into play as is described in ▶ Chap. 65, "Multiscale Concepts in Simulations of Organic Semiconductors," by D. Andrienko. Organic electronics, and more specifically polymer-based electronics, play an increasingly important role in several areas of technology and research such as in the study of organic solar cells, organic light-emitting diodes (OLEDs), or field-effect transistors. The idea is that using soft matter, eventually printable electronics can be developed. For this conjugated organic compounds have to be developed, which not only provide the appropriate intramolecular electronic properties but also form morphologies (amorphous, liquid crystalline, or crystalline), which foster charge transport along the backbones but equally important also between different molecules. Like for inorganic semiconductors, doping is needed as well. Here processing becomes especially important. For these complicated systems, a rational has to be developed to relate device properties to chemical structure and processing history. This cannot be accomplished by simulations alone but requires a close interaction with experiment. The chapter explains the workflow of methodologies to be applied and illustrates this for several examples.

So far examples have been presented, where multiscale modeling is employed in a sequential way, i.e., full systems are studied on different levels of resolution. There are however situations where this is not advisable. In many cases, local chemical detail is needed in some region of space, e.g., docking of a ligand to a protein, while further away details are not needed, and a more coarse description is sufficient to account for, e.g., thermal fluctuations. L. Delle Site and coworkers describe the ▶ Chap. 66, "Adaptive Resolution Molecular Dynamics Technique" which allows the simultaneous treatment of different resolutions in a liquid within a single simulation setup. The class of closely related methods, named AdResS (adaptive resolution simulation) methods, allows the simultaneous simulation of a system on an all-atom, coarse-grained, and continuum level, while the exchange of molecules between different regions of space is not at all restricted. The scheme allows to couple rather different system, e.g., also the coupling of all-atom simulations to an ideal gas, which makes true grand canonical MD simulations easily possible. After describing the general theoretical framework, a few specific examples are given, which illustrate how local chemistry-specific aspects and more global properties influence each other.

So far techniques and concepts have been discussed, which are based on physical concepts and chemical information about the systems. In parallel computer science over the years has developed (deep) learning methods, which are entirely based on available data, i.e., so-called training sets. These methods turn out to be quite successful in many areas, when it comes to the identification/description of general properties. However the inverse design question, namely, to derive a potential energy function from data or to design new chemicals based on "big data" remains a formidable problem. At this point in recent years, a combination of data-driven approaches together with physical/chemical insight has been developed, which leads to a new area of materials simulation and materials discovery. Especially for soft matter, this connection to simulation runs is needed, as entropy is as relevant as energy, unlike for hard matter. T. Bereau introduces these new methodologies in ▶ Chap. 67, "Data-Driven Methods in Multiscale Modeling of Soft Matter." There he also illustrates alternative approaches of sampling in chemical space rather than being confined to physical space of only one kind of system.

All the above chapters deal with soft materials, where thermal fluctuations lead to spatial and temporal averaging or where the response to external driving forces is studied. System evolution beyond thermal fluctuations is triggered by well-defined external sources. There is, however, a rather huge class of soft matter, whose behavior is governed by internal activity, such as bacteria, sperm, colloidal microswimmers, and so on. They consume energy, which can be supplied by various means and develop very interesting motion and structure patterns, especially of interest in biological context. From a modeling point of view, this poses special challenges as one is dealing low Reynolds number hydrodynamics of active systems, which only rather recently also got into the focus of experiment. R.G. Winkler and G. Gompper tackle this problem in their contribution (▶ Chap. 68, "Hydrodynamics in Motile Active Matter"), where several advanced methods to include hydrodynamics effects in dissipative systems are introduced and applied to different classes of microswimmers. The applications range from the study of individual active microswimmers to collective behavior of larger flocks of swimmers.

## 3    Summary/Outlook

The contributions of this general chapter provide an introduction into the very wide field of multiscale polymer/soft matter simulations and should guide to more special and extensive literature (see, e.g., following reviews and books: Frenkel and Smit 2001; Voth 2008; Peter and Kremer 2009; Noid 2013; Potestio et al. 2014; Schiller et al. 2018). Studies of large biomolecular systems as proteins or assemblies of proteins are not covered but are closely connected. The same holds for field-based approaches used in an (mainly commercial) engineering context, such as finite element-related methods (Zheng et al. 2008; Meijer and Govaert 2003). I am sure that the described materials simulations in connection with new advanced data-driven method as well as with links to biomolecular simulations and quantum chemistry on one side and engineering with molecular insight on

the other side will significantly grow and become even more important. Advanced simulation methodologies in general and for soft matter in particular are at a point of transition from reproducing and supporting experiment and theory under well-defined and thus controlled conditions toward true predictive power for materials properties and guidance of experiments. This can lead to new molecular systems, as suggested by advanced sampling in chemical space or to new materials based on molecular insight into specific processing conditions, which only can be acquired by modern simulations approaches. Thus materials discovery by advanced multiscale modeling seems in reach. This software-based development will be further seconded by increasingly available high-performance computing (HPC) power.

# References

De Gennes PG (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca

Doi M (2013) Soft matter physics. Oxford University Press, Oxford

Doi M, Edwards SF (1986) The theory of polymer dynamics. Clarendon Press, Oxford

Frenkel D, Smit B (2001) Understanding molecular simulation, 2nd edn. Academic, San Diego

Meijer HEH, Govaert LE (2003) Multi-scale analysis of mechanical properties of amorphous polymer systems. Macrom Chem Phys 204:274–288

Noid WG (2013) Perspective: coarse-grained models for biomolecular systems. J Chem Phys 139:090901

Peter C, Kremer K (2009) Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back. Soft Matter 5:4357–4366

Potestio R, Peter C, Kremer K (2014) Computer simulations of soft matter: linking the scales. Entropy 16:4199–41245

Rubinstein M, Colby RH (2003) Polymer physics. Oxford University Press, Oxford

Schiller UD, Kruger T, Henrich O (2018) Mesoscopic modelling and simulation of soft matter. Soft Matter 14:9–26

Voth GA (ed) (2008) Coarse-graining of condensed phase and biomolecular systems. CRC Press, Boca Raton

Zheng QH, Yu AB, Lu GQ (2008) Multiscale modeling and simulation of polymer nanocomposites. Prog Polym Sci 33:191–269

# Polymer Solutions

# 61

Burkhard Dünweg

## Contents

### Abstract

The chapter provides a brief general introduction into the concepts of scaling, universality, and crossover scaling, plus the blob concept that provides an intuitive picture of crossover phenomena. We present the most important static and dynamic scaling laws for unentangled uncharged polymer solutions, together with their test and refinement by careful computer simulations. A hoard of simulation methods has been developed for these systems, and these will be briefly discussed as well.

## 1 Introduction

Polymers exist in a variety of states and situations. They may appear as bulk systems or in confined geometries (like films), and they may be solid (semicrystalline,

B. Dünweg (✉)
Theory of Polymers, Max Planck Institute for Polymer Research, Mainz, Germany
Department of Chemical Engineering, Monash University, Clayton, VIC, Australia
e-mail: duenweg@mpip-mainz.mpg.de

rubbery, or glassy) or liquid. In the liquid state, the macromolecules may form a dense melt, or they are dissolved in a solvent of good or poor quality. Finally, for a solution, it makes a huge difference if the molecules are charged or not; in the former case, one has a polyelectrolyte system.

Theoretical physics is mainly interested in the *universal* properties of polymer systems, i.e., those properties that do not depend on the details of the chemistry that defines the monomer. Lots of insight has been gained by deliberately discarding these details and rather focusing on the effects that result from the physics of macromolecules as such. The most important aspects here are (i) connectivity, i.e., the macromolecular architecture, which may be linear chains, rings, stars, combs, networks, etc. (in other words, the topology arising from bonded interactions), (ii) nonbonded interactions (here in particular the excluded-volume interaction, solvent quality effects, and possibly long-range electrostatics), (iii) (possibly) geometric restrictions, and (iv) (for solutions) the dynamic coupling between the motion of macromolecules and the flow of the surrounding solvent (the so-called "hydrodynamic interaction"). This is altogether the huge field of theoretical polymer physics, for which excellent textbooks (De Gennes 1979; Doi and Edwards 1988; Grosberg and Khokhlov 1994; Rubinstein and Colby 2003) exist. Even though the complicated chemistry has been replaced by simplified or coarse-grained models, understanding the physics is still a challenging and complicated problem, which one cannot simply "solve" by straightforward pencil-and-paper analytical theory. Rather, one has to rely on a combination of intuitive insight, theoretical approximations, experiments in the laboratory, as well as careful numerical studies of well-defined models.

The most successful computer models in the "universal" regime of phenomena are (i) simple lattice models and (ii) bead-spring models in the continuum. Both types of models can faithfully model connectivity and interactions (aspects (i) and (ii) of the previous paragraph); however, they have different strengths and weaknesses when it comes to further aspects. Lattice models are particularly well-suited for Monte Carlo studies of static properties, while bead-spring models are particularly amenable for studying the dynamics of systems with hydrodynamic interactions and for applying molecular dynamics and similar methods, which are much easier to parallelize than Monte Carlo algorithms.

The present chapter attempts to provide some overview of the physics (statics and Brownian dynamics) of *neutral* polymer solutions in the *bulk* and computer simulations that have provided confirmation and/or refinement of the underlying ideas. For polyelectrolyte solutions and dense melts, please see the contributions by C. Holm and G. S. Grest in this volume, respectively.

## 2    Scaling Laws

A good deal of theoretical polymer physics is concerned with so-called "scaling laws." This is a very general concept, which has proven extremely useful not only for polymers but also, e.g., in the theory of critical phenomena or the study of

turbulence. From an abstract point of view, scaling laws are nothing but a special
case of the general observation that the mathematical structure of a physical theory is
strongly restricted or perhaps even fully determined by the underlying symmetries.
Here we deal with a special symmetry, which is the invariance under the rescaling of
parameters. Suppose we consider a certain physical quantity $Q$ and we are interested
in its dependence on another physical quantity $P$. As an example, let us think about
the dependence of the average size $R$ of a polymer coil on the contour length $L$
of the (linear) molecule. Let us further assume that we pick a certain value of $P$,
$P = p$, as the basic unit for $P$. Let $q$ be the value of $Q$ for this particular $P$ value.
Then we may write the relation in dimensionless form:

$$\frac{Q}{q} = F\left(\frac{P}{p}\right), \tag{1}$$

where $F$ is a dimensionless function with $F(1) = 1$. Of course, we could also
use a different unit system, by picking a different value $p'$ for the $P$ units and the
corresponding value $q'$ for the $Q$ units, such that we also have

$$\frac{Q}{q'} = \frac{Q}{\lambda q} = G\left(\frac{P}{p'}\right), \tag{2}$$

with (in general) another function $G$, $G(1) = 1$, where the rescaling factor is given
by $\lambda = q'/q$. On the right-hand side, we can similarly introduce $\phi = p'/p$.
Obviously $\phi$ will depend on $\lambda$, $\phi = \phi(\lambda)$. Now, the system is *scale invariant* if
$\phi$ depends *only* on $\lambda$ but not on the point from where the rescaling started (i.e., not
on $p$, $q$) and if also $G = F$. In other words, scale-invariant systems are those that
are characterized by a *lack* of intrinsic scale or those where the physics does *not*
provide some most natural unit system. In our polymer coil example, this means
that it does not matter how long the subchains are into which the full molecule is
decomposed (as long as these subchains are long compared to the size of a chemical
monomer and short compared to the overall contour length). Since we can combine
two rescaling transformations with factors $\lambda$ and $\mu$ into a single one with factor $\lambda\mu$,
we have, for scale-invariant systems,

$$\phi(\lambda\mu) = \phi(\lambda)\phi(\mu) \tag{3}$$

and of course $\phi(1) = 1$. By mapping this relation onto a differential equation, it is
trivial to show that its solution is a power law:

$$\phi(\lambda) = \lambda^{1/\alpha}, \tag{4}$$

with an undetermined exponent $\alpha$. Insertion into Eq. 2 yields

$$\frac{Q}{\lambda q} = F\left(\frac{P}{\lambda^{1/\alpha} p}\right), \tag{5}$$

or, with $x = P/p$,

$$F(x) = \lambda F(\lambda^{-1/\alpha} x). \tag{6}$$

By picking the special value $\lambda = x^{\alpha}$, we see that $F$ is also a power law:

$$F(x) = x^{\alpha}. \tag{7}$$

In other words, scale invariance automatically gives rise to power laws. For our polymer coil example, this means

$$\frac{R}{R_0} = \left(\frac{L}{L_0}\right)^{\nu}, \tag{8}$$

where $R_0$ is the coil size for contour length $L = L_0$ and the Flory exponent $\nu$ depends on the physical situation. For flexible (i.e., sufficiently long) polymers in three dimensions, $\nu$ takes the value $1/2$ for random-walk (RW) statistics, while it is roughly 0.59 for self-avoiding walk (SAW) statistics, which applies in good solvent conditions. Finally, for a chain that forms a collapsed globule due to attractive interactions, $\nu = 1/3$.

It is important to realize that the exponents of scaling laws are typically *universal* (and this is certainly true for $\nu$). This is so because scale invariance means that the system "looks the same" after proper rescaling. Now, the idea of the renormalization group for polymers (Des Cloizeaux and Jannink 1991; Schäfer 1999) is that one should start from an original system and then subject it to a coarse-graining procedure, where several original monomers are lumped into new effective monomers. Iterating this, the chain more and more "forgets" its chemical details, while only the asymptotic scale-invariant structure remains – and this is the same for all original systems within a so-called "universality class." For polymers, all chains with relevant excluded-volume interactions belong to the universality class of SAWs while those with turned-off excluded volume to the RW universality class. For simulations, the concept of universality implies that any model can in principle be used, as long as it falls into the universality class that one wishes to study. This in turn means that the construction of models is mainly guided by considerations of conceptual simplicity, computational efficiency, and convenience in general.

Further important universal quantities are amplitude ratios and crossover scaling functions. The latter will be discussed in the next section; the former are simply the ratios of prefactors of scaling laws in dimensionless form. For example, one can study various measures of the size of a polymer coil, i.e., the end-to-end-distance $R_E \equiv \langle R_E^2 \rangle^{1/2}$, the gyration radius $R_G \equiv \langle R_G^2 \rangle^{1/2}$, and the hydrodynamic radius $R_H \equiv \langle R_H^{-1} \rangle^{-1}$, with

$$R_G^2 = N^{-1} \sum_i (\mathbf{r}_i - \mathbf{R}_{CM})^2, \tag{9}$$

$$\mathbf{R}_{CM} = N^{-1} \sum_i \mathbf{r}_i, \tag{10}$$

$$\mathbf{R}_E = \mathbf{r}_N - \mathbf{r}_1, \tag{11}$$

$$R_H^{-1} = N^{-2} \sum_{i \neq j} |\mathbf{r}_i - \mathbf{r}_j|^{-1}. \tag{12}$$

Here $N$ is the number of monomers of the chain, whose coordinates are denoted with $\mathbf{r}_i$. In the asymptotic long-chain limit, the ratios $R_E/R_G$ and $R_G/R_H$ are universal numbers, taking the values $\sqrt{6}$ and $8/(3\sqrt{\pi})$ for three-dimensional RWs.

## 3    Crossover Scaling

There are many situations where one needs to consider the dependence of a quantity on more than a single variable. For example, in polymer solutions, one is interested in the dependence of the coil size on degree of polymerization $N$, concentration $c$ (total number of monomers per unit volume), and solvent quality. In such a situation, scale invariance is expressed by a straightforward generalization of Eq. 6:

$$F(x_1, x_2, \ldots) = \lambda F(\lambda^{-1/\alpha_1} x_1, \lambda^{-1/\alpha_2} x_2, \ldots). \tag{13}$$

A particularly important case occurs if there are just two arguments, in which case we have

$$F(x_1, x_2) = x_1^{\alpha_1} F(1, x_1^{-\alpha_1/\alpha_2} x_2), \tag{14}$$

such that apart from the power law $x_1^{\alpha_1}$, we also have a dependence on the "crossover scaling variable" $x_c \equiv x_1^{-\alpha_1/\alpha_2} x_2$, while $F(1, x_1^{-\alpha_1/\alpha_2} x_2)$ is then called a "crossover scaling function." Typically the behavior becomes simple in the asymptotic limits $x_c \gg 1$, $x_c \ll 1$, where simple power laws $F \propto x_1^{\beta_>}$, $F \propto x_1^{\beta_<}$ are recovered. In such a case, the crossover scaling function must itself asymptotically behave like an appropriate power law. Finally, if $F$ describes the behavior of a universal ratio, it must itself be universal in the limit of long chains.

## 4    Blobs

Crossover phenomena in polymer solutions can conveniently be described in terms of so-called "blobs." A blob is a portion of the polymer chain that is composed of $g$ monomers and has a typical extension (the "blob size") $\xi$. This length scale marks

the crossover between two different behaviors, and typically an energy of $k_B T$ (thermal energy) is associated with it. The blob concept provides a nice pictorial description of crossover phenomena and is hence a very useful tool for deriving crossover scaling laws. The most important crossovers in the statics of polymer solutions are those between RW and SAW behavior, driven by (i) attractive effective interactions and (ii) concentration.

For a single isolated chain, the quality of solvent can be measured in terms of an effective interaction energy $\epsilon(T)$, which measures the temperature-dependent degree of attraction between two monomers. At the temperature of the theta transition (Lifshitz et al. 1978; Schäfer 1999) ($T = \Theta$), the repulsive and attractive parts of the interaction cancel out, such that effectively the chain behaves as a RW. In the vicinity of $T = \Theta$, we may write $\epsilon(T) = \epsilon_0(1 - \Theta/T)$, which gives rise to a dimensionless interaction parameter $z^\star = (k_B T)^{-1}\epsilon_0(1 - \Theta/T)$. For any $z^\star > 0$, the chain structure is asymptotically a SAW. However, if $z^\star$ is small, the amount of repulsion is too small to disturb the RW statistics on small length scales. This gives rise to a thermal blob size $\xi_T$ corresponding to $g_T$ monomers, of which each has a size $b$, such that $\xi_T \sim b g_T^{1/2}$. The number of monomer-monomer contacts within the blob is estimated as $g_T^{1/2}$. The blob size is found by equating the total energy in the blob with $k_B T$, i.e., $g_T^{1/2}z^\star \sim 1$, or $\xi_T \sim b/z^\star$. Visualizing a very long chain as a SAW composed of RW blobs, one finds $R \sim \xi_T (N/g_T)^\nu$ or

$$R \sim bN^{1/2} \left(N^{1/2}z^\star\right)^{2\nu-1},\tag{15}$$

from which the relevant crossover scaling variable $z = N^{1/2}z^\star$ is read off. For chains that violate the condition $N \gg g_T$, Eq. 15 is generalized to

$$R \sim bN^{1/2}f(z),\tag{16}$$

where the crossover scaling function $f(z)$ behaves like $f(z) \sim z^{2\nu-1}$ for $z \gg 1$ while $f(z) \sim 1$ for $z \ll 1$.

Next, let us consider the concentration-driven crossover from SAW to RW behavior in perfectly good solvent, as a result of Flory screening of excluded-volume interactions (De Gennes 1979; Rubinstein and Colby 2003). The overlap concentration $c^\star$ is the concentration where an arrangement of perfectly swollen chains is just space-filling, i.e., $c^\star \sim N/R^3 \sim b^{-3}N^{-(3\nu-1)}$. At concentrations $c$ that significantly exceed $c^\star$, the chains overlap. This gives rise to concentration blobs of size $\xi_c$, containing $g_c$ monomers. Since on length scales below $\xi_c$ there is no overlap, the SAW structure is unperturbed in this regime. Conversely, on scales above $\xi_c$, we have RW behavior. Therefore, $\xi_c \sim b g_c^\nu$ and $c \sim b^{-3}g_c^{-(3\nu-1)}$ or $\xi_c \sim b(b^3c)^{-\nu/(3\nu-1)}$. A long chain is then a RW composed of SAW blobs, $R \sim \xi_c(N/g_c)^{1/2}$ or

$$R \sim bN^{\nu} \left( \frac{c}{c^{\star}} \right)^{-\frac{\nu-1/2}{3\nu-1}} , \qquad (17)$$

from which the natural crossover scaling variable $c/c^{\star}$ is read off. The generalization in terms of a crossover scaling function is $R \sim bN^{\nu} f(c/c^{\star})$, where now $f(c/c^{\star}) \sim 1$ for $c/c^{\star} \ll 1$, while for $c/c^{\star} \gg 1$ the power law Eq. 17 is recovered. A solution whose monomer concentration is small but whose chains are so long that there is nevertheless a strong overlap is called "semidilute." The semidilute regime ends at a concentration $c^{\star\star}$ where the blob size has shrunken to the monomer size, such that no SAW regime is left.

In the general case, the structure is determined by the interplay between concentration and solvent quality effects or the competition between $\xi_T$ and $\xi_c$. For the ideal picture of a semidilute solution, we have $b \ll \xi_T \ll \xi_c \ll R$. On length scales below $\xi_T$ and above $\xi_c$, we have RW behavior. This is due to attraction on the small scales and due to Flory screening on the large scales. For length scales between $\xi_T$ and $\xi_c$, we have SAW behavior; this regime shrinks more and more upon deteriorating the solvent quality or upon increasing the concentration. This picture gives rise to the generic phase diagram shown in Fig. 1. For more details on the derivation, see the textbook literature or the supplemental material of Jain et al. (2012a). It is important to note that all universal ratios can, in the asymptotic limit of infinite chain length, be expressed in terms of just the two crossover scaling variables $z$ and $c/c^{\star}$.

## 5    Dynamic Scaling

Polymer statics provides us with two important length scales, the monomer size $b$ and the coil size $R$. In case the system needs to be described in terms of a blob picture, there may be a blob size, or possibly even more blob sizes, as additional important length scales. In a dense melt of long chains, there may also be entanglements, which give rise to a "tube diameter" as yet another important length scale. In the present chapter, we will only consider non-entangled systems, where the chains are either too short or too dilute to develop entanglements.

The idea of dynamic scaling for the Brownian motion of polymers may then be understood as follows: For each length $l$, there is an associated time $\tau(l)$. This time may be viewed as the time that a subchain of extension $l$ (in real space, *not* along the contour) needs to move diffusively by its own size. Alternatively, we may also pick a single monomer and study the time dependence of its root mean square displacement, $\langle (\Delta \mathbf{r})^2 \rangle^{1/2}$. The time $\tau(l)$ would then be given by the time that passes until $\langle (\Delta \mathbf{r})^2 \rangle^{1/2} = l$. The underlying concept is here that the single-monomer motion must be consistent with the motion of the object as a whole. Now suppose we consider $l$ values that are well within a regime that is bounded by two of the important scales mentioned in the previous paragraph, with no further important
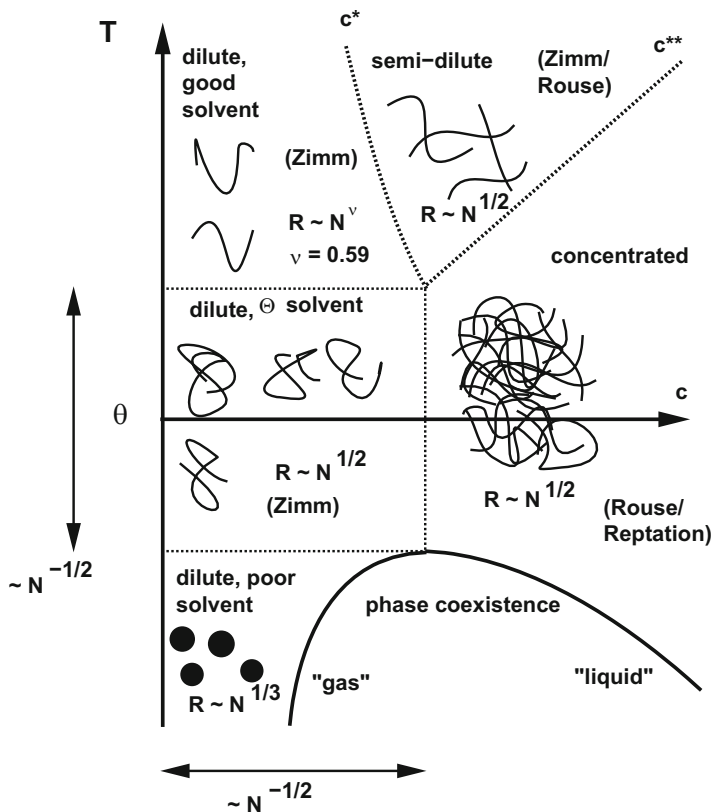
**Fig. 1** Generic phase diagram of a polymer solution

scale in between. Then we again have a lack of a naturally provided unit system, and this applies not only to lengths but also to times. Therefore, the dynamics in such a regime must be described by a power law

$$\tau(l) \propto l^z, \tag{18}$$

which defines the *dynamic exponent z* (not to be confused with the crossover scaling variable of the previous section!), applicable to the regime under consideration.

Let us first consider a system where no blobs occur. This can either be a dilute solution ($c \ll c^\star$) or a dense melt ($c \gg c^{\star\star}$). The longest relaxation time $\tau_R$ is associated with the coil size:

$$\tau_R \propto R^z. \tag{19}$$

Dynamic scaling then implies that the mean square displacement for times $t$ with $\tau(b) \ll t \ll \tau_R$ must obey

$$\left\langle (\Delta \mathbf{r})^2 \right\rangle \propto t^{2/z}. \tag{20}$$

Furthermore, if $D$ denotes the center-of-mass diffusion constant of the chain as a whole, we have $D\tau_R \sim R^2$ or $D \propto R^{-(z-2)}$.

For an isolated chain in solvent (dilute limit), the intra-chain hydrodynamic interaction (i.e., the strong coupling of the monomer motion to the flow of the surrounding solvent) gives rise to so-called Zimm dynamics (Doi and Edwards 1988). A characteristic feature of Zimm dynamics is that the diffusive behavior of the chain is essentially that of a Stokes sphere, $D \sim (k_B T)/(\eta R)$, where $\eta$ is the solvent viscosity. More precisely, the approximate Kirkwood theory (Doi and Edwards 1988) predicts $D = (k_B T)/(6\pi \eta R_H)$. From this, one reads off $z = 3$.

In case the chain is in a dense melt, the hydrodynamic interactions are screened (more about this below). Therefore there are no hydrodynamic correlations in the monomer displacements, and if entanglements play no role (which is the case if the chains are not too long), then the relevant theory for the dynamics is the Rouse model (Doi and Edwards 1988). Here one simply assumes a homogeneous viscous background, and each monomer has a friction constant $\zeta$. The friction coefficients of these monomers simply add up to yield the total friction coefficient of the chain. Therefore $D = (k_B T)/(N\zeta)$. Since $R \sim bN^{1/2}$ (in a dense melt, we have RW statistics), $D \propto R^{-2}$ or $z = 4$.

In a semidilute solution, the length scale that governs the crossover from Zimm to Rouse dynamics is again the blob size $\xi_c$. Up to this length scale, one has unscreened excluded-volume and hydrodynamic interactions, i.e., SAW statistics and Zimm dynamics. The corresponding time scale is the blob relaxation time $\tau_\xi \sim \eta \xi_c^3/(k_B T)$. The blob as a whole has a Stokes friction coefficient $\sim \eta \xi_c$. Computer simulations (Ahlrichs et al. 2001) have shown that hydrodynamic interactions are unscreened as long as the *time* scale is significantly below $\tau_\xi$, regardless of length scales. This is reasonable, since on these short time scales, all monomers just move with the flow, such that correlations exist even on length scales significantly beyond $\xi_c$. Conversely, for times significantly above $\tau_\xi$, the blobs "feel" that they are not free to move (or that chain-chain collisions occur). Therefore, the blobs then dampen the hydrodynamic flow velocity $\mathbf{u}$ with a friction force per unit volume of (De Gennes 1976) $\sim \eta \xi_c \xi_c^{-3}\mathbf{u}$, which is a term that should be added to the Stokes equation for $\mathbf{u}$. For such a situation, the flow field generated by a point force does no longer decay like $1/r$ ($r$: distance from the point force) but rather like $\exp(-r/\xi_H)/r$, where the hydrodynamic screening length $\xi_H \sim \xi_c$. Hydrodynamic screening is thus understood as a randomization of hydrodynamic correlations, induced in essence by chain-chain collisions.

It should also be mentioned that melts do not exhibit strict Rouse dynamics in the dense limit, even if the chains are quite short. The reason is dynamic coupling of the chain motion to the viscoelastic modes of the "matrix," which gives rise to subtle corrections (Farago et al. 2012a, b). Similarly, there are also subtle corrections to the RW statistics of polymer chains in a melt (Wittmer et al. 2004). Both results have been obtained by careful computer simulations.

# 6    Simulations I: Statics

To study universal static single-chain properties, the method of choice is clearly
Monte Carlo (MC) of lattice models, where chains are simply walks on a lattice.
Mostly simple cubic lattices are studied, but other lattice structures are permitted
as well. The so-called pivot algorithm (Madras and Sokal 1988) is presently the
most efficient method known. Here one randomly selects a subchain and rotates
it by a random angle around a random axis. This is an MC trial move, which
is accepted or rejected by the standard Metropolis criterion. Presently the fastest
known implementation is that by N. Clisby (2010), where the information about
the chain conformation is stored in a somewhat unconventional manner: Firstly, one
defines a bounding box about the chain as a whole. Associated with it are global
properties like number of monomers, end-to-end vector, gyration radius, and center-
of-mass coordinate. Then one subdivides the chain into two subchains, for each of
which the analogous information is stored. This is done recursively all the way to
the monomer level, such that one obtains a binary tree. The coordinates of each
box and its contents are stored *relative* to the coordinates of the coarser level. This
makes it possible to move one such "container" as a whole without the need to ever
touch the data of the finer levels that it contains. Similarly, overlap checks are done
by checking the overlap between bounding boxes: If they do not overlap, then their
contents will surely not overlap either. With such tricks, it is possible to reduce the
computational complexity of one pivot move to $\ln N$, such that very long chains are
accessible. A recent study (Clisby and Dünweg 2016) has thus been able to find
for three-dimensional SAWs: $\nu = 0.58759700(40)$, $R_G/R_H = 1.5803940(45)$, and
$R_E^2/R_G^2 = 6.253531(10)$.

Similarly, accurate MC calculations have been able to study the $\Theta$ transition
in three dimensions (Grassberger 1997) and in particular investigate the subtle
logarithmic corrections to scaling that occur there. For this study, the so-called
"PERM" (prune-enriched Rosenbluth method) algorithm was used, where chains
are grown step by step and statistical criteria decide at each step if a chain is
terminated, continued, or even branched to generate yet another chain. In this way,
unbiased samples of long chains may be generated. Universal crossover scaling
functions for the $\Theta$ transition were studied as well (Kumar and Prakash 2003),
using the methodology of Brownian dynamics (BD; we will discuss this method
briefly below). This study emphasized the importance of appropriate extrapolation
procedures: In order to find the crossover scaling function, one should work at a
constant value of the crossover scaling variable (here $z = N^{1/2} z^\star$) and study the
residual dependence of a universal ratio (like $R_G(T)/R_G(T = \Theta)$) on the chain
length. This residual dependence is a correction to scaling; therefore the asymptotic
universal behavior is obtained after extrapolation $N \to \infty$.

The crossover scaling for Flory screening in good solvent was studied utilizing a
lattice model and MC simulation (Paul et al. 1991). However, here the model was the
so-called bond fluctuation model (Carmesin and Kremer 1988), where monomers do
not occupy single sites but rather elementary cubes, while the connecting bonds may
vary within limits. This allows to implement an MC dynamics that involves simply

a random displacement of an elementary cube on the lattice. Indeed it was found that a crossover from SAW to RW statistics occurs, with a crossover length scale $\xi_c$ that exhibits the concentration dependence predicted by blob theory. Later, the same model was also used to reveal the corrections to purely Gaussian behavior in a melt (Wittmer et al. 2004).

Finally, the double crossover that results from the competition between the theta blobs and the Flory screening blobs has recently been studied by BD (Jain et al. 2012a). The internal blob structure could not be resolved, for lack of sufficiently long chains; however the dependence of total chain properties like the coil size on concentration and solvent quality was in perfect agreement with blob theory. The same was true for dynamic properties like the diffusion constant.

# 7   Simulations II: Dynamics

For studies of dynamics, one needs algorithms that faithfully reproduce the motion of the monomers, at least on the (typically long) time scales that one is interested in. Obviously, brute-force molecular dynamics (MD) will satisfy this condition, if it involves all particles in the system. This approach has been highly successful for the studies of melts (see contribution by G. S. Grest in this volume) and can in principle also be applied to the dynamics of solutions, where the hydrodynamic interactions are taken into account by explicit solvent particles. Zimm dynamics for a single chain could thus be successfully established (Pierleoni and Ryckaert 1992; Dünweg and Kremer 1993). For such studies of a single macromolecule in a simulation box, one has to take into account that the latter is typically not much larger than the chain itself. Therefore, one has to deal with strong finite-size effects, which scale as $R_G/L$, where $L$ is the linear box size. This is a direct consequence of the long-range nature of the hydrodynamic interactions: Since the flow field generated by a point force decays like $1/r$ ($r$: distance from the force center), the correlations between the stochastic displacements of two distinct monomers are proportional to the inverse interparticle distance. The theory thus provides detailed quantitative predictions about the magnitude of such finite-size effects, and this in turn makes it possible to quantitatively check Zimm theory even in a finite-box situation.

However, for solutions MD is nearly always unnecessarily expensive and can rather be replaced by cheaper algorithms that simulate the solvent degrees of freedom in a simplified fashion. The only situations where this is not true are either fairly concentrated solutions, where the solvent contribution to the computational effort is only moderate, or investigations of local atomistic dynamics, where local packing and similar phenomena are of specific interest. In all other cases, the effect of the solvent can be summarized by (i) its quality, which may be modeled by just a suitable effective monomer-monomer interaction, and (ii) the hydrodynamic interactions, which give rise to dynamic correlations between monomers as a result of momentum transport through the solvent. The crucial observation is here that the solute-solvent system is characterized by a large separation of time scales: The slowest degree of freedom in the solvent is diffusive momentum transport,

characterized by the "kinematic viscosity" $\eta_{kin} = \eta/\rho$, i.e., the ratio between shear viscosity and mass density, which has the dimension of a diffusion constant. The dimensionless "Schmidt number" $Sc = \eta_{kin}/D$ then relates this to the diffusion constant $D$ of an immersed particle or, more generally, to the diffusion constant of some immersed soft-matter object of size $R$. Typically, in dense fluids $Sc \gg 1$ even for solvent particles, due to a sizeable viscosity value-note $\eta_{kin} \propto \eta$ but $D \propto 1/\eta$. For large (isolated) macromolecules, the corresponding Schmidt number is yet much larger, as $Sc \propto 1/D \propto R$. For these reasons, we may *either* replace the solvent by some sort of "generalized hydrodynamics solver," i.e., a set of more or less artificial degrees of freedom that exhibit the correct hydrodynamic behavior on large length and time scales, *or* dispose of the solvent altogether, by assuming that the flow field follows the configuration of monomers instantaneously, such that it becomes completely enslaved to the latter and thus no longer appears as an explicit degree of freedom.

Let us begin with the latter approach. This is the realm of *Brownian dynamics* (Öttinger 1995). Here one solves a discretized stochastic differential equation for the monomer coordinates $\mathbf{r}_i$, using a finite time step $h$. The update rule can then be written as

$$r_{i\alpha}(t + h) = r_{i\alpha}(t) + h \sum_j \mu_{i\alpha, j\beta} F_{j\beta} + k_B T h \sum_j \partial \mu_{i\alpha, j\beta}/\partial r_{j\beta}$$

$$+ \sqrt{2k_B T h} \sum_j \sigma_{i\alpha, j\beta} q_{j\beta}. \tag{21}$$

Here Greek letters indicate Cartesian indexes with Einstein summation convention. $\mathbf{F}_j$ is the force acting on particle $j$, while $\overset{\leftrightarrow}{\mu}_{ij}$ is the mobility tensor that describes the hydrodynamic correlations between the monomers $i$ and $j$. Typically, the Rotne-Prager tensor (Öttinger 1995) is used. In case one is not interested in correct solution dynamics, one may simply turn the hydrodynamic interactions off and replace $\overset{\leftrightarrow}{\mu}_{ij}$ with a multiple of the unit tensor. In this case, the method will produce Rouse-like dynamics for a single-chain simulation. The last term of Eq. 21 denotes the stochastic displacements, where $q_{i\alpha}$ are random variables with

$$\langle q_{i\alpha} \rangle = 0 \tag{22}$$

$$\langle q_{i\alpha} q_{j\beta} \rangle = \delta_{ij} \delta_{\alpha\beta}, \tag{23}$$

while the matrix $\sigma_{i\alpha, j\beta}$ satisfies

$$\sum_k \sigma_{i\alpha, k\gamma} \sigma_{j\beta, k\gamma} = \mu_{i\alpha, j\beta}. \tag{24}$$

This approach was pioneered by a seminal paper nearly 40 years ago (Ermak and McCammon 1978) and has seen many refinements since then. The main

difficulties are the evaluation of the mobility tensor, which couples all monomers in the system, and the calculation of a suitable square root. These problems have remained a computational challenge for decades. Standard Ewald sums for multichain systems (Jain et al. 2012b) have met moderate success, but only recently has a method been published (Fiore et al. 2017) whose computational effort scales strictly *linearly* with the number of involved monomers.

The generalized hydrodynamics solvers are technically much easier and also much more easy to parallelize. They also scale linearly with the number of monomers, however at the expense of an additional large set of explicit solvent degrees of freedom. These solvers all include thermal fluctuations in some way or another. This is necessary because in soft-matter physics, we are dealing with length and time scales that are so small that fluctuations play a role. Obviously, Brownian motion of polymer chains could not be studied if fluctuations were absent. Therefore such methods are not fully macroscopic but are rather frequently called "mesoscale" methods.

One can distinguish two classes of mesoscale methods, depending on the way how thermal fluctuations are treated. The first class, which one may call "MD-like," are particle methods where the amount of thermal fluctuations per degree of freedom is similar to what one would get in an MD simulation. Peculiar to these methods is the impossibility to adjust the degree of thermal fluctuations *independently* of the macroscopic fluid properties that are relevant for hydrodynamics. Conversely, in the second class, which one may call "hydrodynamics-like," the degree of thermal fluctuations *can* be adjusted independently of the macroscopic properties. The degree of thermal fluctuations is here a reflection of the degree of coarse-graining: The more atomistic particles are lumped into one mesoscale degree of freedom, the smaller is the amount of thermal fluctuations per mesoscale degree of freedom – simply as a result of Gaussian statistics and the law of large numbers.

As this aspect is typically underemphasized in the literature, let us illustrate this by a very simple example, a one-dimensional ideal gas, which we simulate by MD, augmented by a Lowe-Andersen thermostat (Lowe 1999) to bring the system to thermal equilibrium. This thermostat simply picks, from time to time, a pair of nearby particles at random. The center-of-mass velocity of that pair then remains unchanged, while the relative velocity is chosen at random, using the appropriate equilibrium Maxwell-Boltzmann distribution, such that the total momentum is conserved. The thermal (root mean square) velocity of a particle is then $(k_B T/m)^{1/2}$, where $m$ is its mass. This has macroscopic relevance, since this is also the speed of sound. Now let us assume that we lump $M$ adjacent particles into a new mesoscale particle. The new system is then again an ideal gas, which we wish to simulate with the same method. We then have two choices concerning the question of the mass of the mesoscale particles: Either we can assign the value $Mm$, which is naively the correct choice, since the bigger particle should indeed exhibit more inertia. Moreover, the thermal velocity (i.e., the amount of thermal fluctuations) is indeed correctly reduced by the factor $M^{-1/2}$. However, this comes at the price of also reducing the speed of sound by the same factor – and this is a value that we would prefer to keep constant, in order to maintain

the time-scale separation between immersed soft-matter objects and the sound waves. Therefore one typically chooses the value $m$, thus keeping the macroscopic properties intact but overestimating the degree of thermal fluctuations. In other words, MD-like methods are typically too restrictive to permit a fully consistent coarse-graining. This dilemma is solved by the hydrodynamics-like methods, where thermal fluctuations are an add-on with adjustable strength to a method that would also work in the strict macroscopic limit with no fluctuations whatsoever.

Dissipative particle dynamics (DPD) (Espanol and Warren 1995) is directly derived from MD, which is just augmented by a momentum-conserving Langevin thermostat. Similarly to the Lowe-Andersen method, DPD is based upon pairs of nearby particles, which are however not chosen at random but rather considered in their totality, at every single time step. The projection of the relative velocity onto the interparticle axis is dampened by a Langevin friction. This is compensated by stochastic Langevin forces on the two particles that also act along the interparticle axis and add up to zero. The total momentum is conserved, and the fluctuation-dissipation theorem (FDT) is satisfied. A generalized version also thermalizes the velocity components perpendicular to the axis (Junghans et al. 2008); however, it is presently not yet fully understood what effects the implied violation of angular-momentum conservation has on the hydrodynamics.

Quite often, DPD simulations are run with particles that have fairly soft interaction potentials. This is done in the spirit of coarse-graining, which in general leads to such softening of interactions. It also has a practical implication, since softer potentials also allow to use a larger time step. The most radical implementation of that idea is to simply run DPD of an ideal gas as a solvent for soft-matter objects (Smiatek et al. 2008). Using an ideal gas has a huge advantage: The solvent degrees of freedom are reduced to their prime function, which is to transmit momentum through the system, and the equilibrium structure of the immersed objects is unaltered compared to immersion in vacuum. The viscosity can nevertheless be adjusted to reflect dense-fluid conditions, by choosing a sufficiently strong friction.

A yet simpler variant is multiparticle collision dynamics (MPCD) (Gompper et al. 2009). Here the ideal-gas particles are sorted into cubic cells. In each cell, the algorithm determines its local center-of-mass velocity and the relative velocities of the particles with respect to it. The latter are then subjected to a random rotation. This "collision step," which conserves both the momentum and the kinetic energy, serves to thermalize the ideal gas and is followed by a standard MD "streaming step."

Both DPD and MPCD are "MD-like," with a coupling of the monomers to the solvent that arises naturally from the setup of the respective algorithms. We will now turn to the "hydrodynamics-like" methods.

Smoothed dissipative particle dynamics (SDPD) (Espanol and Revenga 2003) has been developed to cure the abovementioned deficiencies of DPD. The name suggests a closer proximity to DPD than the method actually exhibits. While DPD comes in spirit from MD, as essentially a bottom-up approach, SDPD rather is a top-down method: Here the starting point is smoothed particle hydrodynamics

(SPH) (Monaghan 2005), which is nothing but a discretization of the Navier-Stokes equations in terms of particles. This looks deceptively similar to MD but is fundamentally different: Firstly, in MD both the equation of state and also the transport coefficients like the viscosity are an output of the atomistic model and must be determined by simulation. Conversely, in SPH they are input parameters. Secondly, MD particles have as only properties their coordinates and momenta (and possibly their orientations and angular momenta). SPH particles, on the other hand, have additional properties "on board" that one could not even *define* for MD particles because their nature is genuinely *thermodynamic* – volume and entropy, which both change in the course of time as a result of the dynamics. SDPD adds Langevin noise to the SPH equations of motion such that the FDT is satisfied. Although the SDPD particles are thermodynamic objects, it is nevertheless possible to simply connect a set of them via springs and thus obtain an immersed polymer chain with the correct large-scale properties (Litvinov et al. 2008). The polymer-solvent coupling is therefore as straightforward as for DPD and MPCD.

Instead of discretizing the Navier-Stokes equations in terms of particles, one may also discretize them via a lattice. One therefore arrives at standard finite-difference or finite-volume schemes (Donev et al. 2010; Balboa-Usabiaga et al. 2012). Again, one may add thermal fluctuations to the equations to satisfy the FDT.

Finally, one may also simulate hydrodynamics via the Lattice Boltzmann (LB) (Dünweg and Ladd 2009) method. Here one solves a linearized and fully discretized version of the Boltzmann equation known from the kinetic theory of gases. Space and time are discretized in terms of a lattice spacing $a$ and time step $h$, respectively. Velocity space is also discretized and reduced to a small discrete set of velocities $\mathbf{c}_i$. Each lattice site contains a set of real-valued positive variables $n_i$, which are interpreted as the mass density corresponding to velocity $\mathbf{c}_i$. The mass density $\rho$ and the momentum density $\mathbf{j}$ are then obtained as zeroth and first velocity moment of the populations:

$$\rho = \sum_i n_i, \tag{25}$$

$$\mathbf{j} = \sum_i n_i \mathbf{c}_i. \tag{26}$$

The procedure then begins with a collision step, i.e., a rearrangement of the populations on the site such that mass and momentum are conserved:

$$n_i \to n_i^{\star} = n_i + \Delta_i, \tag{27}$$

where the "collision operator" $\Delta_i$ satisfies

$$\sum_i \Delta_i = 0, \tag{28}$$

$$\sum_i \Delta_i \mathbf{c}_i = 0. \tag{29}$$

This is followed by a streaming to the adjacent lattice sites, such that the total procedure can be written in terms of the Lattice Boltzmann equation (LBE):

$$n_i(\mathbf{r} + \mathbf{c}_i h, t + h) = n_i(\mathbf{r}, t) + \Delta_i(\mathbf{r}, t). \tag{30}$$

This implies that the discrete velocities must be chosen commensurate with the lattice. For example, the popular D3Q19 model (Dünweg and Ladd 2009), which lives on the three-dimensional simple cubic lattice, involves 19 velocities, which correspond to the 6 nearest and 12 next-nearest neighbors, plus the zero velocity.

The method involves lots of adjustable parameters, like the set of velocity shells, associated weight coefficients, and various details of the collision operator. All of these are tuned in order to obtain the correct Navier-Stokes behavior in the continuum limit, which is found from the algorithm by subjecting the LBE to an asymptotic (Chapman-Enskog) analysis. The LBE can therefore be used as a Navier-Stokes equation solver in its own right. Thermal fluctuations are introduced by adding a suitably chosen stochastic collision operator to $\Delta_i$. For further details, see Dünweg and Ladd (2009).

It should be emphasized that in all of the abovementioned "hydrodynamics-like" methods, it is very important to make sure that the FDT is not only satisfied in the asymptotic continuum limit but rather for the algorithm as such. Substantial effort has gone into the development of methods that do satisfy this condition.

In contrast to particle methods, hybrid methods that involve MD for the polymer chains and a lattice algorithm for the solvent need special care for the fluid-particle coupling. A particularly simple approach is a frictional coupling (Dünweg and Ladd 2009), where each monomer is assigned a Stokes friction coefficient. Therefore each monomer is not only subject to the conservative forces coming from other monomers (and possibly yet other sources) but also to a friction force and a Langevin stochastic force. The former dampens the *relative* velocity of the particle with respect to the local flow field, which is obtained via interpolation from adjacent lattice sites. The latter is just standard Langevin noise that is needed to satisfy the FDT. Back coupling is obtained by interpolating the thus-resulting momentum transfer back to the lattice and enforcing momentum conservation. Another possibility is to enforce a stick boundary condition, either on the surface of an extended particle  (Dünweg and Ladd 2009) or based upon a point-particle picture (Usabiaga et al. 2013).

At the end of this section, we briefly wish to mention a few studies that have focused on polymer solution dynamics. Zimm dynamics of a single chain has been studied by BD by many authors, e.g., Fixman (1986), Liu and Dünweg (2003), and Sunthar and Prakash (2006), where the last study also investigated the solvent quality-driven crossover behavior. Single-chain Zimm dynamics was also studied by LB/MD (Ahlrichs and Dünweg 1999), MPCD (Mussawisade et al. 2005), and

SDPD (Litvinov et al. 2008). Not surprisingly, all these studies yield essentially the same results, and it is even possible to quantitatively map them onto each other – this has explicitly been done for LB/MD vs. pure MD (Ahlrichs and Dünweg 1999) as well as for LB/MD vs. BD (Pham et al. 2009).

   Detailed studies of the concentration-driven crossover from Zimm to Rouse dynamics have been done by both LB/MD (Ahlrichs et al. 2001) and MPCD (Huang et al. 2010). Both confirmed the picture of hydrodynamic screening as outlined in Sect. 5, and the latter paper went even beyond to also study non-equilibrium behavior.

## 8    Summary

Polymer solution statics and dynamics are beautiful pieces of physics where progress has been made by analytical theory (in particular scaling considerations), experiments, and computer simulations. Improved physical and mathematical insight led to the development of computer simulation methods that went from simple and fairly brute force to more and more sophisticated and problem-oriented, focusing on the essence of the phenomena one wishes to study. The author hopes that the present contribution has given the reader a glimpse on how fruitfully theory and simulations have worked together in this field. For reasons of both space and also expertise of the author, the present chapter has only focused on the most basic equilibrium phenomena and completely left out the highly important field of non-equilibrium physics, i.e., nonlinear polymer solution rheology, which would be worth yet another chapter in this series.

## References

Ahlrichs P, Dünweg B (1999) Simulation of a single polymer chain in solution by combining lattice Boltzmann and molecular dynamics. J Chem Phys 111(17):8225–8239. https://doi.org/10.1063/1.480156

Ahlrichs P, Everaers R, Dünweg B (2001) Screening of hydrodynamic interactions in semidilute polymer solutions: a computer simulation study. Phys Rev E 64(4):040501. https://doi.org/10.1103/PhysRevE.64.040501

Balboa-Usabiaga F, Bell J, Delgado-Buscalioni R, Donev A, Fai T, Griffith B, Peskin C (2012) Staggered schemes for fluctuating hydrodynamics. Multiscale Model Simul 10(4):1369–1408. https://doi.org/10.1137/120864520

Carmesin I, Kremer K (1988) The bond fluctuation method – a new effective algorithm for the dynamics of polymers in all spatial dimensions. Macromolecules 21(9):2819–2823. https://doi.org/10.1021/ma00187a030

Clisby N (2010) Accurate estimate of the critical exponent $\nu$ for self-avoiding walks via a fast implementation of the pivot algorithm. Phys Rev Lett 104(5):055702. https://doi.org/10.1103/PhysRevLett.104.055702

Clisby N, Dünweg B (2016) High-precision estimate of the hydrodynamic radius for self-avoiding walks. Phys Rev E 94(5):052102. https://doi.org/10.1103/PhysRevE.94.052102

De Gennes PG (1976) Dynamics of Entangled polymer solutions. II. Inclusion of hydrodynamic interactions. Macromolecules 9(4):594–598. https://doi.org/10.1021/ma60052a012

De Gennes PG (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca, New York

Des Cloizeaux J, Jannink G (1991) Polymers in solution: their modelling and structure. Clarendon Press, Oxford

Doi M, Edwards SF (1988) The theory of polymer dynamics. Oxford University Press, Oxford

Donev A, Vanden-Eijnden E, Garcia A, Bell J (2010) On the accuracy of finite-volume schemes for fluctuating hydrodynamics. Commun Appl Math Comput Sci 5(2):149–197. https://doi.org/10.2140/camcos.2010.5.149. http://msp.org/camcos/2010/5-2/p01.xhtml

Dünweg B, Kremer K (1993) Molecular dynamics simulation of a polymer chain in solution. J Chem Phys 99(9):6983–6997. https://doi.org/10.1063/1.465445

Dünweg B, Ladd AJC (2009) Lattice Boltzmann simulations of soft matter systems. In: Holm C, Kremer K (eds) Advanced computer simulation approaches for soft matter sciences III. Advances in polymer science, vol 221. Springer, Berlin/Heidelberg, pp 89–166. https://doi.org/10.1007/978-3-540-87706-6_2.

Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. J Chem Phys 69(4):1352–1360. https://doi.org/10.1063/1.436761

Espanol P, Revenga M (2003) Smoothed dissipative particle dynamics. Phys Rev E 67(2):026705. https://doi.org/10.1103/PhysRevE.67.026705

Espanol P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. Europhys Lett 30(4):191. https://doi.org/10.1209/0295-5075/30/4/001. http://stacks.iop.org/0295-5075/30/i=4/a=001

Farago J, Meyer H, Baschnagel J, Semenov AN (2012a) Mode-coupling approach to polymer diffusion in an unentangled melt. II. The effect of viscoelastic hydrodynamic interactions. Phys Rev E 85(5):051807. https://doi.org/10.1103/PhysRevE.85.051807

Farago J, Semenov AN, Meyer H, Wittmer JP, Johner A, Baschnagel J (2012b) Mode-coupling approach to polymer diffusion in an unentangled melt. I. The effect of density fluctuations. Phys Rev E 85(5):051806. https://doi.org/10.1103/PhysRevE.85.051806

Fiore AM, Balboa Usabiaga F, Donev A, Swan JW (2017) Rapid sampling of stochastic displacements in Brownian dynamics simulations. J Chem Phys 146(12):124116. https://doi.org/10.1063/1.4978242

Fixman M (1986) Implicit algorithm for Brownian dynamics of polymers. Macromolecules 19(4):1195–1204. https://doi.org/10.1021/ma00158a042

Gompper G, Ihle T, Kroll DM, Winkler RG (2009) Multi-particle collision dynamics: a particle-based mesoscale simulation approach to the hydrodynamics of complex fluids. In: Holm C, Kremer K (eds) Advanced computer simulation approaches for soft matter sciences III. Advances in polymer science, vol 221. Springer, Berlin/Heidelberg, pp 1–87. https://doi.org/10.1007/12_2008_5

Grassberger P (1997) Pruned-enriched Rosenbluth method: simulations of $\theta$ polymers of chain length up to 1,000,000. Phys Rev E 56(3):3682–3693. https://doi.org/10.1103/PhysRevE.56.3682

Grosberg AY, Khokhlov AR (1994) Statistical physics of macromolecules. AIP Press, New York

Huang CC, Winkler RG, Sutmann G, Gompper G (2010) Semidilute polymer solutions at equilibrium and under shear flow. Macromolecules 43(23):10107–10116. https://doi.org/10.1021/ma101836x

Jain A, Dünweg B, Prakash JR (2012a) Dynamic crossover scaling in polymer solutions. Phys Rev Lett 109(8):088302. https://doi.org/10.1103/PhysRevLett.109.088302

Jain A, Sunthar P, Dünweg B, Prakash JR (2012b) Optimization of a Brownian-dynamics algorithm for semidilute polymer solutions. Phys Rev E 85(6):066703. https://doi.org/10.1103/PhysRevE.85.066703

Junghans C, Praprotnik M, Kremer K (2008) Transport properties controlled by a thermostat: an extended dissipative particle dynamics thermostat. Soft Matter 4(1):156–161. https://doi.org/10.1039/B713568H. http://pubs.rsc.org/en/Content/ArticleLanding/2008/SM/B713568H

Kumar KS, Prakash JR (2003) Equilibrium swelling and universal ratios in dilute polymer solutions: exact Brownian dynamics simulations for a delta function excluded volume potential. Macromolecules 36(20):7842–7856. https://doi.org/10.1021/ma034296f

Lifshitz IM, Grosberg AY, Khokhlov AR (1978) Some problems of the statistical physics of polymer chains with volume interaction. Rev Modern Phys 50(3):683–713. https://doi.org/10.1103/RevModPhys.50.683

Litvinov S, Ellero M, Hu X, Adams NA (2008) Smoothed dissipative particle dynamics model for polymer molecules in suspension. Phys Rev E 77(6):066703. https://doi.org/10.1103/PhysRevE.77.066703

Liu B, Dünweg B (2003) Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation. J Chem Phys 118(17):8061–8072. https://doi.org/10.1063/1.1564047

Lowe CP (1999) An alternative approach to dissipative particle dynamics. Europhys Lett 47(2):145. https://doi.org/10.1209/epl/i1999-00365-x/meta

Madras N, Sokal AD (1988) The pivot algorithm: a highly efficient Monte Carlo method for the self-avoiding walk. J Stat Phys 50(1–2):109–186. https://doi.org/10.1007/BF01022990

Monaghan JJ (2005) Smoothed particle hydrodynamics. Rep Prog Phys 68(8):1703. https://doi.org/10.1088/0034-4885/68/8/R01. http://stacks.iop.org/0034-4885/68/i=8/a=R01

Mussawisade K, Ripoll M, Winkler RG, Gompper G (2005) Dynamics of polymers in a particle-based mesoscopic solvent. J Chem Phys 123(14):144905. https://doi.org/10.1063/1.2041527

Öttinger HC (1995) Stochastic processes in polymeric fluids: tools and examples for developing simulation algorithms. Springer, Berlin/New York

Paul W, Binder K, Heermann DW, Kremer K (1991) Crossover scaling in semidilute polymer solutions: a Monte Carlo test. J Phys II 1(1):37–60. https://doi.org/10.1051/jp2:1991138

Pham TT, Schiller UD, Prakash JR, Dünweg B (2009) Implicit and explicit solvent models for the simulation of a single polymer chain in solution: lattice Boltzmann versus Brownian dynamics. J Chem Phys 131(16):164114. https://doi.org/10.1063/1.3251771

Pierleoni C, Ryckaert J (1992) Molecular dynamics investigation of dynamic scaling for dilute polymer solutions in good solvent conditions. J Chem Phys 96(11):8539–8551. https://doi.org/10.1063/1.462307

Rubinstein M, Colby RH (2003) Polymer physics. Oxford University Press, Oxford/New York

Schäfer L (1999) Excluded volume effects in polymer solutions. Springer, Berlin/Heidelberg. https://doi.org/10.1007/978-3-642-60093-7

Smiatek J, Allen MP, Schmid F (2008) Tunable-slip boundaries for coarse-grained simulations of fluid flow. Eur Phys J E 26(1–2):115–122. https://doi.org/10.1140/epje/i2007-10311-4

Sunthar P, Prakash JR (2006) Dynamic scaling in dilute polymer solutions: the importance of dynamic correlations. Europhys Lett 75(1):77. https://doi.org/10.1209/epl/i2006-10067-y/meta

Usabiaga FB, Pagonabarraga I, Delgado-Buscalioni R (2013) Inertial coupling for point particle fluctuating hydrodynamics. J Comput Phys 235:701–722. https://doi.org/10.1016/j.jcp.2012.10.045. http://www.sciencedirect.com/science/article/pii/S0021999112006493

Wittmer JP, Meyer H, Baschnagel J, Johner A, Obukhov S, Mattioni L, Müller M, Semenov AN (2004) Long range bond-bond correlations in dense polymer solutions. Phys Rev Lett 93(14):147801. https://doi.org/10.1103/PhysRevLett.93.147801

# From the Atomistic to the Macromolecular Scale: Distinct Simulation Approaches for Polyelectrolyte Solutions

# 62

## Jens Smiatek and Christian Holm

## Contents

## Abstract

Polyelectrolytes reveal interesting properties in solution. At short length scales, the dissociation of counterions is heavily affected by the chemical structure of the polyelectrolyte, the properties of the solution, and specific ion effects. At larger length scales, the structure of polyelectrolyte solutions is dominated by long-range interactions. In the special case of dissolved polyanions and polycations, polyelectrolyte complexes or multilayers can form. In this review we present distinct simulation approaches to study the corresponding effects at different

J. Smiatek
Institute for Computational Physics, University of Stuttgart, Stuttgart, Germany
Helmholtz Institute Münster (HI MS), Ionics in Energy Storage, Forschungszentrum Jülich GmbH, Münster, Germany
e-mail: smiatek@icp.uni-stuttgart.de

C. Holm (✉)
Institute for Computational Physics, University of Stuttgart, Stuttgart, Germany
e-mail: holm@icp.uni-stuttgart.de

length scales in more detail. Whereas at short length scales, atomistic molecular dynamics simulation is often the method of choice, semi-coarse-grained and coarse-grained models with a lower level of details reveal their benefits at larger length scales.

# 1    Introduction

In accordance with the IUPAC definition, polyelectrolytes are charged macro-molecules, in which a substantial portion of the constitutional units contains ionic or ionizable groups, or both (McNaught and Wilkinson 1997). Strong polyelectrolytes like DNA are completely ionized, whereas weak polyelectrolytes like polyacrylic acid show an incomplete and pH-dependent dissociation behavior. Consequently, the properties of polyelectrolytes are dominated by a combination of electrostatic and molecular interactions, which heavily affect the occurring conformations and the resulting ion dissociation behavior. In contrast to uncharged polymers, the long-range decay of electrostatic interactions between the ionic groups impedes the derivation of scaling relations in terms of simple mean-field approaches (de Gennes 1979; Doi and Edwards 1988; Dobrynin and Rubinstein 2005; Dobrynin 2008). Thus, the complex interplay between the polyelectrolyte, the ions in the solution, and the solvent reveals many interesting phenomena at different length scales, which can be studied in more detail with the help of distinct molecular dynamics (MD) simulation approaches in combination with appropriate models for the considered species.

At short length scales, previous atomistic MD simulations demonstrated that the molecular solvation behavior significantly influences the amount of ionized groups and thus the corresponding conformation with regard to counterion condensation effects and repulsive electrostatic interactions along the polyelectrolyte backbone (Smiatek et al. 2014). Vice versa, the behavior of the surrounding solvent molecules is also modified by the presence of ionic groups, which is reflected by local varia-tions concerning the dielectric permittivity and the solvation of ions (Fahrenberger et al. 2015a; Mukhopadhyay et al. 2012). In contrast to these local interactions, the aggregation of polyelectrolytes becomes important at larger length scales, such that polycations and polyanions form polyelectrolyte complexes and multilayers, whose stability significantly depends on enthalpic and entropic contributions and the properties of the embedded solvent molecules (Cerdà et al. 2009; Qiao et al. 2011, 2012). Also in bulk solution, molecular properties determine the configurational behavior of the polyelectrolyte, as can be seen by the formation of polyelectrolyte micelles, pearl-necklace structures, or the onset of microphase separation processes between polar and apolar regions (Limbach and Holm 2003; Limbach et al. 2004; Dormidontova et al. 1994). Thus, appropriate models with a sufficient degree of detail are needed for reliable simulations of polyelectrolyte solutions in order to elucidate the properties of interest.

In this chapter, we introduce several well-established approaches for the simu-lation of polyelectrolyte systems at different length and time scales. The benefits

and drawbacks of atomistic, semi-coarse-grained and simple coarse-grained models are discussed, and we will present representative examples for various applications. In the last section, we summarize the main points and address current limitations concerning the general applicability of the methods.

## 2    Simulation Approaches for Polyelectrolyte Solutions at Different Length Scales

The discussion of methods in this section follows a bottom-up approach, which means that we start with atomistic models and the description of molecular force fields. Hereafter, we decrease the level of resolution in terms of semi- and simple coarse-grained approaches, which are best suited for the simulation of processes at large length and long time scales. All models rely on the consideration of explicit particles in combination with molecular dynamics or Monte Carlo time integration schemes. It has to be noted that other continuum or self-consistent field approaches can also be used, for instance, as described in Schmid (1998). Over the last years, multiscale simulation approaches were also developed, which rely on a combined consideration of the following methods in terms of matching and adaptive resolution schemes. The reader is referred to Fritz et al. (2011) and Praprotnik et al. (2008) for more details and for other contributions to this volume.

### 2.1    Atomistic Models: Importance of Chemical and Molecular Details

Due to the massive increase of computational power over the last years, it is nowadays possible to study the properties of short polyelectrolyte chains, the so-called oligoelectrolytes, for hundreds of nanoseconds via atomistic MD simulations in combination with accurate molecular force fields. In terms of modern force fields, molecular properties are dictated by bonded and nonbonded interactions, which are represented by classical potential functions in order to avoid an extremely expensive evaluation of the electronic behavior. In more detail, standard atomistic force fields like OPLS/AA (Jorgensen et al. 1996) include parameters for bonded, angular, and dihedral potentials in addition to nonbonded interactions, which are usually represented by Coulomb and Lennard-Jones contributions. In combination with advanced electrostatics algorithms (Deserno and Holm 1998; Arnold et al. 2013) and the sophisticated use of graphics processing units (GPUs), the properties of polyelectrolyte solutions with a dimension up to several nanometers can be studied for hundreds of nanoseconds by standard atomistic MD approaches. Hence, detailed information on the molecular interactions, the solvent behavior, and the corresponding distribution functions are accessible.

As an illustrative example, the stable conformations for a sulfonated oligosulfonic acid with sodium counterions in water and chloroform are displayed in Fig. 1, where more details of this system can be found in Smiatek et al. (2014). It can
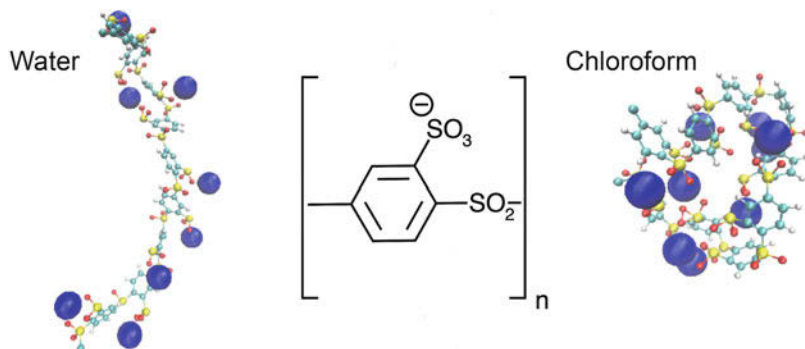
**Fig. 1** Sulfonated oligosulfonic acid with nine monomers according to $N_p = 9$ and an equivalent number of sodium counterions (blue spheres) in water (left side) and chloroform (right side). The solvent molecules are not shown for reasons of clarity, and the chemical structure of one monomer is depicted in the middle

be clearly seen that the oligoelectrolyte reveals a swollen conformation in polar solvents like water when compared with chloroform. This finding can be related to molecular solvation effects and standard polymer theories, which distinguish between good, poor, and theta solvents (de Gennes 1979). Thus, the average size of the polyelectrolyte with $N_p$ monomers can be described by a power-law behavior according to $R \sim N_p^\nu$ with the excluded-volume parameter $\nu$, which is usually $\nu = 0.588$ for uncharged polymers in good solvents (Doi and Edwards 1988). Depending on the actual value of the excluded-volume parameter, good and poor solvents are defined by $\nu > 1/2$ and $\nu < 1/2$, respectively, whereas a theta solvent is characterized by $\nu = 1/2$. A polyelectrolyte chain at infinite dilution should scale with $\nu = 1$. Hence, the molecular properties of the polyelectrolyte adapt directly to the polarity of the solvent, which is reflected by an increase or a decrease of the solvent-accessible surface area and the size.

Interestingly, further analysis concerning atomistic models for polyelectrolytes revealed that the counterion behavior differs significantly from standard theoretical descriptions (Manning 1969, 1996; Deserno et al. 2000; Deserno and Holm 2001), which highlights the benefits of atomistic MD approaches in order to verify molecular theories and to study the corresponding deviations. As a result of these simulations (Lund et al. 2008; Heyda and Dzubiella 2012; Smiatek et al. 2014; Batys et al. 2017), it was found that the counterion distribution around the polyelectrolyte can deviate from standard mean-field predictions in terms of the Poisson-Boltzmann theory (Andelman 1995), which can be attributed to specific solvation effects and molecular interactions (Smiatek et al. 2014). Furthermore, it was also observed that specific ion effects (Marcus 2009; Kunz 2010; Lo Nostro and Ninham 2012) and conformational properties of polyelectrolytes (Wohlfarth et al. 2015) influence the corresponding counterion distribution and yield results in good agreement with experimental data. To summarize, also atomistic simulations reveal short-range deviations from simple mean-field theories; however, Poisson-Boltzmann theory

and its generalizations often agree surprisingly well for global observables like the osmotic pressure or effects that are dominated by the Coulomb interactions (Deserno et al. 2001; Antypov et al. 2005; Antypov and Holm 2006; Lu et al. 2008).

## 2.2    Semi-Coarse-Grained Approaches: MARTINI Force Fields and Refined Models

Atomistic MD simulations mainly intend to study the properties of short oligoelectrolyte chains and their molecular interactions with counterions and other components of the solution. Although also larger systems were studied, for instance, polyelectrolyte complexes and multilayers (Farhat et al. 1999; Micciulla et al. 2014), it has to be mentioned that these simulations are very CPU-time demanding and usually restricted to several tens of nanoseconds, such that even a reasonable equilibration of these complex solutions is hard to achieve. In order to introduce efficient but still accurate models for these larger systems, the use of semi-coarse-grained approaches is highly beneficial. In general, coarse-grained approaches intend to reduce the degrees of freedom when compared with atomistic simulations, which results in a significant acceleration of the computation times and the possibility of using larger time steps (Winger et al. 2009; Marrink et al. 2010). Standard coarse-grained frameworks for polyelectrolytes and polymers include the introduction of simple bead-spring models (Doi and Edwards 1988). Hence, several atoms of a monomer are represented by one interaction site, the so-called bead, while the individual beads are connected with springs, which are usually modeled by classic harmonic or modified Finitely Extensible Nonlinear Elastic (FENE) potentials. In contrast to the most simple coarse-grained methods, semi-coarse-grained approaches like the MARTINI force field (Marrink and Tieleman 2013) refine these very generic models by consideration of important chemical details. For instance, the MARTINI force field usually relies on a 4:1 mapping scheme, such that four heavy atoms are combined into one interaction site (CG bead) with parameterized polarity values and hydrogen bond acceptor and donor abilities (Marrink et al. 2007). In more detail, the MARTINI CG beads can be divided into different particle-type classes (polar (P), nonpolar (N), apolar (C), and charged (Q) species). The subtypes within these classes are categorized due to their ability to form hydrogen bonds (donor (d), acceptor (a), both donor and acceptor (da), and none of them (0)) and with regard to their polarity (from 1 = low polarity to 5 = high polarity) (Marrink et al. 2007; Marrink and Tieleman 2013). Further subclasses were also defined to increase the local resolution (Marrink et al. 2007). All CG beads reveal different Lennard-Jones parameters and partial charges, which were parameterized according to partitioning coefficients for similar atomic groups in oil/water mixtures (Marrink et al. 2007; Marrink and Tieleman 2013).

Over the last years, several MARTINI models for polymers and polyelectrolytes were developed. Specific examples are DNA (Uusitalo et al. 2015), poly(styrene sulfonate) (PSS) and poly(diallyldimethylammonium) (PDADMA), which were both used for the study of polyelectrolyte complexes (Vögele et al. 2015a). The
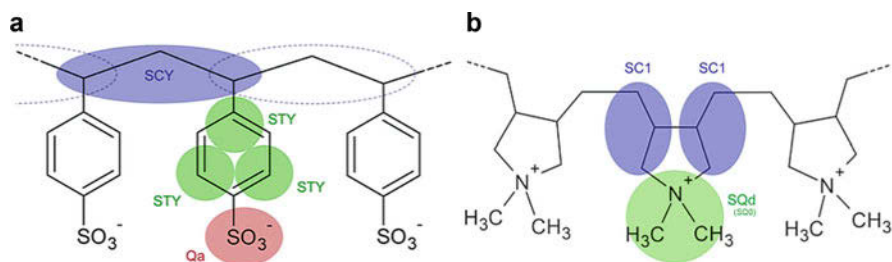
**Fig. 2** Chemical structure of PSS (left side) and PDADMA (right side) with three monomers. The differently shaded regions denote spherical MARTINI CG beads with particle-type classes as defined in Marrink et al. (2007)

molecular topologies of PSS and PDADMA in terms of the MARTINI force field and the corresponding particle types according to Marrink et al. (2007) are shown in Fig. 2. In combination with refined polarizable water models (Yesylevskyy et al. 2010; Michalowsky et al. 2017), it was shown that MARTINI simulations are well suited to study the properties of highly charged systems at large length and time scales (Vögele et al. 2015a, b; Uusitalo et al. 2015).

The advantages of MARTINI models are mostly given by their flexibility and their transferability, but it has to be noted that solvent particles indeed have to be considered explicitly. In order to circumvent time-consuming calculations, the matching of potential of mean forces between atomistic and coarse-grained simulations provides a computationally efficient route in terms of tabulated potentials and thus an implicit solvent approach (Brini et al. 2013; Reith et al. 2002; Li et al. 2012; Lyubartsev and Laaksonen 1999; Savelyev and Papoian 2010; Hsu et al. 2012). Although this method, which is also often called iterative Boltzmann inversion technique (Reith et al. 2003), avoids time-consuming calculation of interactions between solvent particles and between solvent and polyelectrolyte groups, it has to be mentioned that the force-matching method is mostly applicable for homogeneous solutions without interfaces. More refined coarse-grained models for DNA and ionomer systems, based on the matching of ion mobilities, were published in Lu et al. (2014), Weik et al. (2016), and Rau et al. (2017). A semi-coarse-grained approach for the simulation of weak polyelectrolytes was recently introduced in Landsgesell et al. (2017a, b). Furthermore, the well-known decrease of the dielectric permittivity around charged objects and the corresponding consequences were recently studied in coarse-grained polyelectrolyte solution via a modification of the Maxwell equation molecular dynamics algorithm (Fahrenberger and Holm 2014; Fahrenberger et al. 2015a, b).

In summary, refined or semi-coarse-grained models can be used for the study of systems at intermediate length and time scales. Nevertheless, for the study of long time-scale processes like transport behavior or the influence of hydrodynamics on polyelectrolyte motion, the use of simple coarse-grained methods remains the most suitable choice.

## 2.3    Simple Coarse-Graining: Generic Bead-Spring Models with Explicit Charges

In general, all simple coarse-grained models are composed of single interaction sites, which have a lower resolution when compared with semi-coarse-grained approaches and thus usually correspond to individual monomers or the number of monomers within the corresponding persistence length (Doi and Edwards 1988). All adjacent beads are connected by springs in terms of simple harmonic or FENE potentials, which restrict the length of the bonds to the equilibrium distance in order to avoid entanglement effects in polymer melts (Kremer and Grest 1990; Stevens and Kremer 1993a, b). In contrast to semi-coarse-grained approaches, simple bead-spring models do not include angular or dihedral potentials per definition, and the corresponding nonbonded interactions are represented by simple Lennard-Jones (LJ) and Coulomb interactions. Most often, the WCA potential, a purely repulsive shifted and truncated version of the Lennard-Jones potential, is used to mimic hard spheres for beads and ions (Weeks et al. 1971). In contrast to MARTINI models, the solvent is often modeled implicitly by consideration of a global dielectric constant, which is thus inserted into the Coulomb potential. Furthermore, the model can be even more simplified by using a screened electrostatic potential or neglecting Coulomb interactions all together (Hickey et al. 2012; Szuttor et al. 2017; Roy et al. 2017). Specifically in solvents with high values of the dielectric constant, for instance, water, electrostatic interactions between charged groups and ions dominate only at short distances (Collins 2004). This can be mostly attributed to the low value of the Bjerrum length at room temperature $\lambda_B = e^2/4\pi\epsilon_0\epsilon_r k_B T$ with the elementary charge $e$, the vacuum permittivity $\epsilon_0$, the dielectric constant $\epsilon_r$, the temperature $T$, and the Boltzmann constant $k_B$. The Bjerrum length estimates the distance where the thermal energy dominates over the electrostatic energy and which is for water at 300 K around $\lambda_B \approx 0.7$ nm corresponding to two hydration shells (Collins 2004; Marcus and Hefter 2006). For larger and highly charged objects like polyelectrolytes or colloidal particles, also the salt concentration plays a significant role in order to induce a fast decay of electrostatic interactions. Hence, by a simple linearization of the Poisson-Boltzmann equation in terms of the Debye-Hückel approximation, the corresponding electrostatic screening length reads $\lambda_D = (\epsilon_r\epsilon_0 k_B T/(\sum_i 2z_i^2 e^2\rho_i))^{1/2}$ with the valency $z_i$ and the ion density $\rho_i$ (Andelman 1995). With regard to this relation, one usually obtains a screening length of $\lambda_D \approx 1$ nm for water at room temperature with a salt concentration of 0.1 mol/L, which implies that the monomers are only weakly affected by the electric field of the surrounding polyelectrolyte groups (Szuttor et al. 2017). Hence, for large objects like λ-DNA, it is often sufficient to neglect electrostatic interactions between the monomers, if the Debye and the Bjerrum length are significantly smaller than the size of the polyelectrolyte in accordance with $\lambda_D \approx \lambda_B \ll R$ (Szuttor et al. 2017; Roy et al. 2017).

Most of these simple coarse-grained models are used to study the dynamics of polyelectrolytes and other components in solution, often under the influence of

external forces in order to induce transport processes. Recent reviews highlighted the benefits of these models in combination with sophisticated mesoscopic simulation techniques for the study of transport processes and electrokinetic effects in microchannels (Slater et al. 2009; Pagonabarraga et al. 2010; Smiatek et al. 2012). In more detail, mesoscopic simulation approaches induce a stochastic motion of the solute species in the system, which is related to the behavior at long time scales. A simple but effective and thermodynamically consistent approach is represented by Langevin or Brownian dynamics (Kremer and Grest 1990). However, as it has been pointed out in Dünweg (1993), momentum is not conserved in Langevin dynamics, such that this approach cannot capture any hydrodynamical effects (Doi and Edwards 1988; Ober and Thomas 1997; Grass et al. 2008; Frank and Winkler 2009). Most often, one is indeed specifically interested in hydrodynamic effects, such that either the Langevin dynamics approach can be modified by introducing the Oseen tensor (Ermak and McCammon 1978) or efficient Navier-Stokes solvers have to be used. In terms of mesoscopic approaches, the most common techniques are dissipative particle dynamics (DPD) (Groot and Warren 1997; Smiatek et al. 2012), coupled lattice Boltzmann/molecular dynamics (LBMD) (Dünweg and Ladd 2009) or multiparticle collision dynamics (MPCD) (Gompper et al. 2008). Previous articles highlighted the good quantitative agreement between DPD and LBMD simulations (Smiatek et al. 2009) and their applicability to study transport processes in confined geometries (Smiatek and Schmid 2010, 2011; Smiatek et al. 2012; Weik et al. 2016). In general, the use of mesoscopic simulation methods can be nowadays regarded as a standard approach in order to study the influence of hydrodynamics in many different research fields.

Furthermore, simple coarse-grained models were often used to validate analytical mean-field approaches like counterion condensation theories (Deserno et al. 2000), stretching forces on tethered polymers (Szuttor et al. 2017), electrohydrodynamic screening effects (Grass et al. 2008), or the combined influence of electroosmotic and electrophoretic motion (Smiatek and Schmid 2010). Interestingly, also the influence of the solvent quality on the resulting conformations can be studied via the use of simple approaches. In terms of experimental and atomistic simulation results, it is known that poor solvents imply collapsed conformations, whereas good solvents lead to a swelling of the polyelectrolyte. Hence, these findings can be transferred to tunable attractive or repulsive interactions between the beads in order to correct for the presence of an implicit solvent. Hence, in combination with electrostatic interactions and attractive bead potentials, it was even possible to enforce the occurrence of pearl-necklace polyelectrolyte structures (Dobrynin et al. 1996; Micka et al. 1999; Limbach et al. 2002; Limbach and Holm 2003). Furthermore, also more complicated topologies like in polymeric ionic liquids (PILs) (Mecerreyes 2011; Yuan et al. 2013) can be modeled via simple coarse-grained approaches (Weyman et al. 2018). A simple example for a common alkylimidazolium-based PIL and a snapshot of a system conformation are shown in Fig. 3. As can be seen in the bottom, a simulation of 30 PIL chains with $N_p = 30$ reveals the occurrence of a microphase separation between polar and apolar beads. The aggregation of apolar beads was initiated by attractive LJ interactions, which initiate the formation

**Fig. 3** Chemical structure of a typical alkylimidazolium-based PIL with counterions X$^-$ (top left). The corresponding simple bead-spring model with $N_p = 4$ is shown at the upper right part. The charged terminal groups are shown as red spheres and the counterions as blue spheres. All neutral beads have a gray color. A snapshot of the simulation with 30 polymers with $N_p = 30$ in terms of the coarse-grained model is shown at the bottom. Neutral beads are colored in red, while charged beads and counterions are colored in green
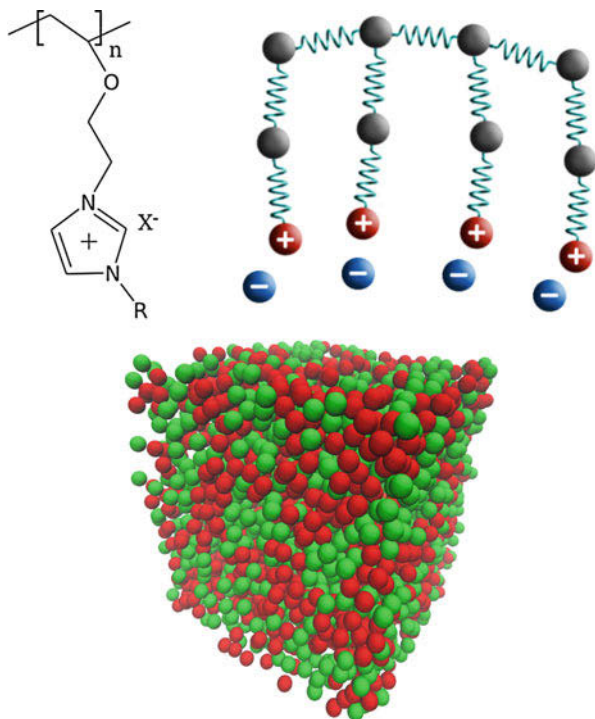


of apolar and polar microphases for larger side chain lengths. In summary, simple coarse-grained polyelectrolyte models are computationally efficient and can be used for the study of different systems at larger scales. The consideration of further details can be simply achieved via effective generic or tunable potentials and provides reasonable results, if the specific molecular details and processes are of minor importance.

## 3   Summary and Conclusion

In this chapter, we described distinct approaches for the simulation of polyelectrolyte solutions at different length and time scales. For the study of fast processes and molecular interactions at short length scales, the use of atomistic models with regard to appropriate force fields is advised. In terms of larger length and longer time scales, semi- and simple coarse-grained models with different levels of detail can often be considered as the method of choice. A promising approach is the MARTINI force field, which provides a beneficial transferability between different systems without the need of a proper reparameterization for distinct models. Furthermore, also tabulated potential methods as well as refined coarse-grained models were also developed that can be interpreted as coarse-grained approaches with basic molecular

properties. For the simulation of transport processes and hydrodynamic effects, simple coarse-grained models in combination with mesoscopic simulation methods are highly beneficial, which can be rationalized by the fact that polymers show a universal scaling behavior at large scales, such that specific molecular details are of minor importance (de Gennes 1979; Doi and Edwards 1988).

In summary, the presented methods can be used for a broad range of systems at different time and length scales. However, which method is most appropriate will depend crucially on the question to be answered, and one cannot give any general advice. However, it has to be emphasized that all methods rely on crucial approximations, and thus, if possible, one should always verify the simulation results with experimental findings or theoretical predictions. A prominent example are specific ion effects, which are modeled explicitly only in few atomistic MD force fields (Fyta and Netz 2012). It is thus a challenging task to transfer this information to coarse-grained models. Moreover, the study of apolar organic solvents is significantly more complicated in comparison to polar solvents, due to the fact that the Debye and the Bjerrum lengths can easily exceed the simulation box size, which induces electrostatic correlation effects between the polyelectrolyte and its periodic images. These finite-size effects modify structural and dynamic properties of the solution and are thus a crucial problem in order to bring simulation outcomes in quantitative agreement with experimental results. Moreover, also non-ideal effects, as they are well known for higher component solutions (Krishnamoorthy et al. 2016), are often not correctly reproduced. In general, all considered methods rely on potential functions to mimic the electronic behavior. Hence, an accurate study of bond formation and cleavage processes in combination with varying pH values of the solution is often impossible. A promising new route is the introduction of reactive force fields (Senftle et al. 2016), which are, although time consuming, less computationally expensive than ab initio simulations. In accordance with the simplification of the electronic behavior, atomic polarization effects are also usually neglected, which have nowadays become an active field of research and model improvement (Lemkul et al. 2016; Bordin et al. 2016).

Despite their limitations, the presented methods and models are the most promising approaches for the reliable study of effects and processes in polyelectrolyte solutions. It can be expected that new and refined approaches in combination with longer simulation times will allow a more accurate study of these systems in the coming years.

# References

Andelman D (1995) Electrostatic properties of membranes: the Poisson-Boltzmann theory, in Handbook of biological physics. North-Holland pp. 603–642

Antypov D, Holm C (2006) Optimal cell approach to osmotic properties of finite stiff-chain polyelectrolytes. Phys Rev Lett 96:088302

Antypov D, Barbosa MC, Holm C (2005) A simple non-local approach to treat size correlations within Poisson-Boltzmann theory. Phys Rev E 71:061106

Arnold A, Fahrenberger F, Holm C, Lenz O, Bolten M, Dachsel H, Halver R, Kabadshow I, Gähler F, Heber F, Iseringhausen J, Hofmann M, Pippig M, Potts D, Sutmann G (2013) Comparison of scalable fast methods for long-range interactions. Phys Rev E 88:063308. https://doi.org/10.1103/PhysRevE.88.063308

Batys P, Luukkonen S, Sammalkorpi M (2017) Ability of Poisson–Boltzmann equation to capture molecular dynamics predicted ion distribution around polyelectrolytes. Phys Chem Chem Phys 19:24583–24593

Bordin JR, Podgornik R, Holm C (2016) Static polarizability effects on counterion distributions near charged dielectric surfaces: a coarse-grained molecular dynamics study employing the drude model. Eur Phys J Special Top 225(8):1693–1705. https://doi.org/10.1140/epjst/e2016-60150-1

Brini E, Algaer EA, Ganguly P, Li C, Rodriguez-Ropero F, van der Vegt NFA (2013) Systematic coarse-graining methods for soft matter simulations – a review. Soft Matter 9:2108–2119. https://doi.org/10.1039/C2SM27201F

Cerdà JJ, Qiao B, Holm C (2009) Understanding polyelectrolyte multilayers: an open challenge for simulations. Soft Matter 5:4412–4425. https://doi.org/10.1039/b912800j

Collins KD (2004) Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. Methods 34(3):300–311. https://doi.org/10.1016/j.ymeth.2004.03.021

de Gennes PG (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca. http://books.google.com/books?id=ApzfJ2LYwGUC&lpg=PP1&num=15&pg=PP1#v=onepage&q&f=false

Deserno M, Holm C (1998) How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. J Chem Phys 109:7678

Deserno M, Holm C (2001) Cell-model and poisson-boltzmann-theory: a brief introduction. In: Holm C, Kékicheff P, Podgornik R (eds) Electrostatic effects in soft matter and biophysics, NATO science series II – mathematics, physics and chemistry, vol 46. Kluwer Academic Publishers, Dordrecht, pp 27–50

Deserno M, Holm C, May S (2000) Fraction of condensed counterions around a charged rod: comparison of Poisson-Boltzmann theory and computer simulations. Macromolecules 33:199–206. https://doi.org/10.1021/ma990897o

Deserno M, Holm C, Blaul J, Ballauff M, Rehahn M (2001) The osmotic coefficient of rod-like polyelectrolytes: computer simulation, analytical theory, and experiment. Eur Phys J E 5:97–103

Dobrynin AV (2008) Theory and simulations of charged polymers: from solution properties to polymer nanomaterials. Curr Opin Colloid Interface Sci 13:376–388. https://doi.org/10.1016/j.cocis.2008.03.006, http://www.sciencedirect.com/science/article/pii/S1359029408000411

Dobrynin AV, Rubinstein M (2005) Theory of polyelectrolytes in solutions and at surfaces. Prog Polym Sci 30(11):1049–1118. https://doi.org/10.1016/j.progpolymsci.2005.07.006, http://www.sciencedirect.com/science/article/B6TX2-4H2G8WN-1/2/4b378f3016fcf641ad1821b4ded9d389

Dobrynin AV, Rubinstein M, Obukhov SP (1996) Cascade of transitions of polyelectrolytes in poor solvents. Macromolecules 29(8):2974

Doi M, Edwards SF (1988) The theory of polymer dynamics. Oxford University Press, Oxford

Dormidontova EE, Erukhimovich IY, Khokhlov AR (1994) Microphase separation in poor-solvent polyelectrolyte solutions: phase diagram. Macromol Theory Simul 3(4):661–675

Dünweg B (1993) Molecular dynamics algorithms and hydrodynamic screening. J Chem Phys 99(9):6977–82

Dünweg B, Ladd AJC (2009) Lattice boltzmann simulations of soft matter systems. In: Advanced computer simulation approaches for soft matter sciences III. Advances in polymer science, vol 221. Springer, Berlin, pp 89–166. https://doi.org/10.1007/12_2008_4

Ermak DL, McCammon J (1978) Brownian dynamics with hydrodynamic interactions. J Chem Phys 69:1352

Fahrenberger F, Holm C (2014) Computing the Coulomb interaction in inhomogeneous dielectric media via a local electrostatics lattice algorithm. Phys Rev E 90:063304. https://doi.org/10.1103/PhysRevE.90.063304

Fahrenberger F, Hickey OA, Smiatek J, Holm C (2015a) Importance of varying permittivity on the conductivity of polyelectrolyte solutions. Phys Rev Lett 115:118301. http://link.aps.org/doi/10.1103/PhysRevLett.115.118301

Fahrenberger F, Hickey OA, Smiatek J, Holm C (2015b) The influence of charged-induced variations in the local permittivity on the static and dynamic properties of polyelectrolyte solutions. J Chem Phys 143:243140. http://scitation.aip.org/content/aip/journal/jcp/143/24/10.1063/1.4936666

Farhat T, Yassin G, Dubas ST, Schlenoff JB (1999) Water and ion pairing in polyelectrolyte multilayers. Langmuir 15(20):6621–6623

Frank S, Winkler RG (2009) Mesoscale hydrodynamic simulation of short polyelectrolytes in electric fields. J Chem Phys 131(23):234905. https://doi.org/10.1063/1.3274681

Fritz D, Koschke K, Harmandaris VA, van der Vegt NF, Kremer K (2011) Multiscale modeling of soft matter: scaling of dynamics. Phys Chem Chem Phys 13(22):10412–10420

Fyta M, Netz RR (2012) Ionic force field optimization based on single-ion and ion-pair solvation properties: going beyond standard mixing rules. J Chem Phys 136(12):12410–12413. https://doi.org/10.1063/1.3693330

Gompper G, Ihle T, Kroll DM, Winkler RG (2008) Multi-particle collision dynamics: a particle-based mesoscale simulation approach to the hydrodynamics of complex fluids. Adv Polym Sci 221:1–87

Grass K, Böhme U, Scheler U, Cottet H, Holm C (2008) Importance of hydrodynamic shielding for the dynamic behavior of short polyelectrolyte chains. Phys Rev Lett 100:096104

Groot RD, Warren PB (1997) Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. J Chem Phys 107(11):4423–4435

Heyda J, Dzubiella J (2012) Ion-specific counterion condensation on charged peptides: Poisson–Boltzmann vs. atomistic simulations. Soft Matter 8(36):9338–9344

Hickey OA, Shendruk TN, Harden JL, Slater GW (2012) Simulations of free-solution electrophoresis of polyelectrolytes with a finite debye length using the Debye-Hückel approximation. Phys Rev Lett 109:098302. https://doi.org/10.1103/PhysRevLett.109.098302

Hsu CW, Fyta M, Lakatos G, Melchionna S, Kaxiras E (2012) Ab initio determination of coarse-grained interactions in double-stranded DNA. J Chem Phys 137(10):105102

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236

Kremer K, Grest GS (1990) Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. J Chem Phys 92(8):5057–5086

Krishnamoorthy AN, Zeman J, Holm C, Smiatek J (2016) Preferential solvation and ion association properties in aqueous dimethyl sulfoxide solutions. Phys Chem Chem Phys 18:31312–31322. https://doi.org/10.1039/C6CP05909K

Kunz W (ed) (2010) Specific ion effects. World Scientific, Singapore

Landsgesell J, Holm C, Smiatek J (2017a) Simulation of weak polyelectrolytes: a comparison between the constant pH and the reaction ensemble method. Eur Phys J Special Top 226:725–736. https://doi.org/10.1140/epjst/e2016-60324-3

Landsgesell J, Holm C, Smiatek J (2017b) Wang-Landau reaction ensemble method: simulation of weak polyelectrolytes and general acid-base reactions. J Chem Theory Comput 13(2):852–862. https://doi.org/10.1021/acs.jctc.6b00791

Lemkul JA, Huang J, Roux B, MacKerell AD Jr (2016) An empirical polarizable force field based on the classical Drude oscillator model: development history and recent applications. Chem Rev 116(9):4983–5013. https://doi.org/10.1021/acs.chemrev.5b00505

Li YC, Wen TC, Wei HH (2012) Electrophoretic stretching of tethered polymer chains by travelling-wave electric fields: tunable stretching, expedited coil–stretch transition, and a new paradigm of dynamic molecular probing. Soft Matter 8(6):1977–1990

Limbach HJ, Holm C (2003) Single-chain properties of polyelectrolytes in poor solvent. J Phys Chem B 107(32):8041–8055

Limbach HJ, Holm C, Kremer K (2002) Structure of polyelectrolytes in poor solvent. Europhys Lett 60(4):566–572

Limbach HJ, Sayar M, Holm C (2004) Polyelectrolyte bundles. J Phys Condens Matter 16(22):2135–2144

Lo Nostro P, Ninham BW (2012) Hofmeister phenomena: an update on ion specificity in biology. Chem Rev 112(4):2286–2322. https://doi.org/10.1021/cr200271j

Lu BZ, Zhou YC, Holst MJ, McCammon JA (2008) Recent progress in numerical methods for the poisson-boltzmann equation in biophysical applications. Commun Comput Phys 3(5): 973–1009

Lu K, Rudzinski JF, Noid W, Milner ST, Maranas JK (2014) Scaling behavior and local structure of ion aggregates in single-ion conductors. Soft Matter 10(7):978–989

Lund M, Vácha R, Jungwirth P (2008) Specific ion binding to macromolecules: effects of hydrophobicity and ion pairing. Langmuir 24(7):3387–3391

Lyubartsev A, Laaksonen A (1999) Effective potentials for ion–DNA interactions. J Chem Phys 111(24):11207–11215

Manning GS (1969) Limiting laws and counterion condensation in polyelectrolyte solutions I. colligative properties. J Chem Phys 51:924–933

Manning GS (1996) Counterion condensation theory constructed from different models. Physica A 231(1–3):236–253

Marcus Y (2009) Effect of ions on the structure of water: structure making and breaking. Chem Rev 109(3):1346–1370. https://doi.org/10.1021/cr8003828

Marcus Y, Hefter G (2006) Ion pairing. Chem Rev 106(11):4585–4621

Marrink SJ, Tieleman DP (2013) Perspective on the MARTINI model. Chem Soc Rev 42(16):6801–6822. https://doi.org/10.1039/C3CS60093A

Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111(27):7812–7824. https://doi.org/10.1021/jp071097f

Marrink SJ, Periole X, Tieleman DP, de Vries AH (2010) Comment on using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models by M. Winger, D. Trzesniak, R. Baron and WF van Gunsteren. Phys Chem Chem Phys 2009, 11, 1934. Phys Chem Chem Phys 12(9):2254–2256

McNaught AD, Wilkinson A (1997) Compendium of chemical terminology, vol 1669. Blackwell Science, Oxford

Mecerreyes D (2011) Polymeric ionic liquids: broadening the properties and applications of polyelectrolytes. Prog Polym Sci 36(12):1629–1648. https://doi.org/10.1016/j.progpolymsci.2011.05.007

Micciulla S, Sanchez PA, Smiatek J, Qiao B, Sega M, Laschewsky A, Holm C, von Klitzing R (2014) Layer-by-layer formation of oligoelectrolyte multilayers: a combined experimental and computational study. Soft Mater 12:S14. https://doi.org/10.1080/1539445X.2014.930046, http://www.tandfonline.com/eprint/eCn9vDIc5aMbyBmB6DV5/full

Michalowsky J, Schäfer LV, Holm C, Smiatek J (2017) A refined polarizable water model for the coarse-grained MARTINI force field with long-range electrostatic interactions. J Chem Phys 146(5):054501. https://doi.org/10.1063/1.4974833

Micka U, Holm C, Kremer K (1999) Strongly charged, flexible polyelectrolytes in poor solvents –
a molecular dynamics study. Langmuir 15:4033

Mukhopadhyay A, Fenley AT, Tolokh IS, Onufriev AV (2012) Charge hydration asymmetry:
the basic principle and how to use it to test and improve water models. J Phys Chem B
116(32):9776–9783

Ober MMCK, Thomas EL (1997) Competing interactions and levels of ordering in self-organizing
polymeric materials. Science 277:1225–1232

Pagonabarraga I, Rotenberg B, Frenkel D (2010) Recent advances in the modelling and simulation
of electrokinetic effects: bridging the gap between atomistic and macroscopic descriptions. Phys
Chem Chem Phys 12:9566–9580. https://doi.org/10.1039/C004012F

Praprotnik M, Junghans C, Site LD, Kremer K (2008) Simulation approaches to soft matter:
generic statistical properties vs. chemical details. Comput Phys Commun 179(1–3):51

Qiao B, Sega M, Holm C (2011) An atomistic study of a poly(styrene sul-
fonate)/poly(diallyldimethylammonium) bilayer: the role of surface properties and charge
reversal. Phys Chem Chem Phys 13(36):16336–16342. https://doi.org/10.1039/C1CP21777A

Qiao B, Sega M, Holm C (2012) Properties of water in the interfacial region of a polyelectrolyte
bilayer adsorbed onto a substrate studied by computer simulations. Phys Chem Chem Phys
14:11425–11432. https://doi.org/10.1039/C2CP41115F

Rau T, Weik F, Holm C (2017) A dsDNA model optimized for electrokinetic applications.
Soft Matter 3918–3926. https://doi.org/10.1039/C7SM00270J, http://pubs.rsc.org/en/content/
articlehtml/2017/sm/c7sm00270j

Reith D, Müller B, Müller-Plathe F, Wiegand S (2002) How does the chain extension of poly
(acrylic acid) scale in aqueous solution? A combined study with light scattering and computer
simulation. J Chem Phys 116(20):9100–9106

Reith D, Pütz M, Müller-Plathe F (2003) Deriving effective mesoscale potentials from atomistic
simulations. J Comput Chem 24(13):1624–1636

Roy T, Szuttor K, Smiatek J, Holm C, Hardt S (2017) Stretching of surface-tethered polymers
in pressure-driven flow under confinement. Soft Matter 13:6189–6196. https://doi.org/10.1039/
C7SM00306D

Savelyev A, Papoian GA (2010) Chemically accurate coarse graining of double-stranded DNA.
Proc Natl Acad Sci U S A 107(47):20340–20345. https://doi.org/10.1073/pnas.1001163107

Schmid F (1998) Self-consistent-field theories for complex fluids. J Phys Condens Matter
10(37):8105

Senftle TP, Hong S, Islam MM, Kylasa SB, Zheng Y, Shi YK, Junkermeier C, Engel-Herbert R,
Janik MJ, Aktulga HM, Verstraelen T, Grama A, van Duin ACT (2016) The ReaxFF reactive
force-field: development, applications and future directions. Comput Mater 2:15011. https://doi.
org/10.1038/npjcompumats.2015.11

Slater GW, Holm C, Chubynsky MV, de Haan HW, Dubé A, Grass K, Hickey OA, Kingsburry C,
Sean D, Shendruk TN, Zhan L (2009) Modeling the separation of macromolecules: a review of
current computer simulation methods. Electrophoresis 30(5):792–818. https://doi.org/10.1002/
elps.200800673

Smiatek J, Schmid F (2010) Polyelectrolyte electrophoresis in nanochannels: a dissipative particle
dynamics simulation. J Phys Chem B 114(19):6266–6272. https://doi.org/10.1021/jp100128p,
http://pubs.acs.org/doi/abs/10.1021/jp100128p

Smiatek J, Schmid F (2011) Mesoscopic simulations of electroosmotic flow and electrophoresis
in nanochannels. Comput Phys Commun 182(9):1941–1944. https://doi.org/10.1016/j.cpc.
2010.11.021, http://www.sciencedirect.com/science/article/pii/S0010465510004674. Computer
Physics Communications Special Edition for Conference on Computational Physics Trondheim,
Norway, 23–26 June 2010

Smiatek J, Sega M, Holm C, Schiller UD, Schmid F (2009) Mesoscopic simulations of the
counterion-induced electro-osmotic flow: a comparative study. J Chem Phys 130:244702.
https://doi.org/10.1063/1.3152844

Smiatek J, Harishchandra RK, Rubner O, Galla HJ, Heuer A (2012) Properties of compatible solutes in aqueous solution. Biophys Chem 160(1):62–68. https://doi.org/10.1016/j.bpc.2011.09.007

Smiatek J, Wohlfarth A, Holm C (2014) The solvation and ion condensation properties for sulfonated polyelectrolytes in different solvents-a computational study. New J Phys 16(2):025001. http://stacks.iop.org/1367-2630/16/i=2/a=025001

Stevens MJ, Kremer K (1993a) Form factor of salt-free linear polyelectrolytes. Macromolecules 26:4717

Stevens MJ, Kremer K (1993b) Structure of salt-free linear polyelectrolytes. Phys Rev Lett 71:2228

Szuttor K, Roy T, Hardt S, Holm C, Smiatek J (2017) The stretching force on a tethered polymer in pressure-driven flow. J Chem Phys 147(3):034902. https://doi.org/10.1063/1.4993619

Uusitalo JJ, Ingólfsson HI, Akhshi P, Tieleman DP, Marrink SJ (2015) MARTINI coarse-grained force field: extension to DNA. J Chem Theory Comput 11(8):3932–3945

Vögele M, Holm C, Smiatek J (2015a) Coarse-grained simulations of polyelectrolyte complexes: MARTINI models for poly(styrene sulfonate) and poly(diallyldimethylammonium). J Chem Phys 143:243151. https://doi.org/10.1063/1.4937805, http://scitation.aip.org/content/aip/journal/jcp/143/24/10.1063/1.4937805

Vögele M, Holm C, Smiatek J (2015b) Properties of the polarizable MARTINI water model: a comparative study for aqueous electrolyte solutions. J Mol Liq 212:103–110. https://doi.org/10.1016/j.molliq.2015.08.062, http://www.sciencedirect.com/science/article/pii/S0167732215304657

Weeks JD, Chandler D, Andersen HC (1971) Role of repulsive forces in determining the equilibrium structure of simple liquids. J Chem Phys 54:5237

Weik F, Kesselheim S, Holm C (2016) A coarse-grained DNA model for the prediction of current signals in DNA translocation experiments. J Chem Phys 145(19):194106. https://doi.org/10.1063/1.4967458

Weyman A, Bier M, Holm C, Smiatek J (2018) Microphase separation and the formation of ion conductivity channels in poly (ionic liquid) s: a coarse-grained molecular dynamics study. J Chem Phys 148:193824

Winger M, Trzesniak D, Baron R, van Gunsteren WF (2009) On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. Phys Chem Chem Phys 11(12):1934–1941

Wohlfarth A, Smiatek J, Kreuer KD, Takamuku S, Jannasch P, Maier J (2015) Proton dissociation of sulfonated polysulfones: influence of molecular structure and conformation. Macromolecules 48(4):1134–1143. https://doi.org/10.1021/ma502550f

Yesylevskyy SO, Schäfer LV, Sengupta D, Marrink SJ (2010) Polarizable water model for the coarse-grained MARTINI force field. PLoS Comput Biol 6(6):e1000810. https://doi.org/10.1371/journal.pcbi.1000810

Yuan J, Mecerreyes D, Antonietti M (2013) Poly(ionic liquid)s: an update. Prog Polym Sci 38:1009–1036. https://doi.org/10.1016/j.progpolymsci.2013.04.002

# Resolving Properties of Entangled Polymers Melts Through Atomistic Derived Coarse-Grained Models

# 63

Gary S. Grest, K. Michael Salerno, Brandon L. Peters, Ting Ge, and Dvora Perahia

## Contents

### Abstract

Coupled length and time scales determine the dynamic behavior of polymers and polymer nanocomposites, thus causing their unique properties. To resolve the properties over large time and length scales it is imperative to develop coarse-grained models which retain atomistic specificity. Here we probe the degree of coarse graining required to access large length and time scales and simultaneously retain significant atomistic details. The degree of coarse graining

G. S. Grest (✉) · B. L. Peters
Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, USA
e-mail: gsgrest@sandia.gov; blpete@sandia.gov

K. Michael Salerno
Center for Materials Physics and Technology, US Naval Research Laboratory, Washington, DC, USA
e-mail: ksalerno@pha.jhu.edu

T. Ge
Department of Mechanical Engineering and Materials Science, Duke University, Durham, USA
e-mail: ting.ge@duke.edu

D. Perahia
Department of Chemistry and Department of Physics and Astronomy, Clemson University, Clemson, SC, USA
e-mail: dperahi@g.clemson.edu

in turn sets the minimum length scale instrumental in defining polymer properties and dynamics. Using polyethylene as a model system, we probe how the scale of coarse graining affects the measured dynamics with different number of methylene groups per coarse-grained bead. Using these models, it is currently possible to simulate polyethylene melts for times of order 1 millisecond. This allows one to study a wide range of properties from chain mobility to viscoelastic response for well-entangled polymer melts while retaining atomistic detail.

# 1　　Introduction

Polymer properties depend on a wide range of coupled length and time scales, with unique viscoelastic properties stemming from interactions down to the atomistic level. The need to probe polymers across time and length scales to capture polymer behavior makes probing dynamics, and particularly computational modeling, inherently challenging. With increasing molecular weight, polymer melts become highly entangled, and the long-time diffusive regime becomes computationally inaccessible using atomistic simulations. While the largest length scales of polymer dynamics are controlled by entanglements, the shortest time and length scales required to resolve dynamic properties are not obvious. This knowledge is critical for developing models that can transpose atomistic details into the long-time scales needed to model long, entangled polymer chains.

One path to overcoming the computational challenge of large time and length scales in polymers and polymer nanocomposites is to coarse grain (CG) the polymer, reducing the number of degrees of freedom and increasing the fundamental time scale. The effectiveness of this process depends on retaining the smallest length scale essential to capturing the polymer dynamics. The process of coarse graining amounts to combining groups of atoms into pseudoatom beads and determining the bead interaction potentials. Simple models like the bead-spring model (Kremer and Grest 1990; Grest 2016) capture the main characteristics of polymers but disregard atomistic details. These models cannot quantitatively describe properties like structure, local dynamics, or densities. Numerous recent studies have worked to bridge the divide between atomistic and coarse models, developing new approaches to drive computational studies to larger length and time scales while maintaining relevant sub-nanometer details (Müller-Plathe 2002; Peter and Kremer 2009; Fritz et al. 2011; Li et al. 2013). The mapping scheme which defines which atoms are combined into a pseudoatom bead is not unique and depends on the specific system and the local properties one wants to retain. For polystyrene (PS), for example, there are at least seven different mapping schemes in which each PS monomer is represented by either one or two CG beads (Karimi-Varzabeh et al. 2012). For flexible polymers, such as polyethylene (Fukunaga et al. 2002; Padding and Briels 2002; Guerrault et al. 2004; Ashbaugh et al. 2005; Chen et al. 2006; Padding and Briels 2011), polybutadiene (Maurel et al. 2012), poly(dimethylsiloxane), and polyisobutylene (Maurel et al. 2015), multiple monomers are combined to form one CG bead. However, few of these studies have explored the effect of varying the

degree of coarse graining on the properties of polymer melts (Abrams and Kremer 2003; Harmandaris et al. 2007; Karimi-Varzabeh et al. 2012; Salerno et al. 2016a, b; Dallavalle and van der Vegt 2017; Peters et al. 2017).

Here, using linear polyethylene as a model system, we review how the degree of coarse graining affects the macromolecular structure and dynamics (Salerno et al. 2016a, b; Peters et al. 2017). The backbone of polyethylene (PE) consists of -$CH_2$– methylene groups, which are a natural coarse-graining unit. Probably the most well-known CG model for PE is the united atom (UA) model (Siepmann et al. 1993; Paul et al. 1995; Martin and Siepmann 1998; Nath et al. 1998), which combines each $CH_2$/$CH_3$ group into one pseudoatom. The UA interaction parameters are determined phenomenologically to reproduce physical properties such as densities and critical temperatures. Going beyond the UA model, PE has previously been studied using CG models with beads of $\lambda = 3$–$48$ methylene groups per CG bead (Padding and Briels 2001; Fukunaga et al. 2002; Padding and Briels 2002; Guerrault et al. 2004; Ashbaugh et al. 2005; Chen et al. 2006; Curcó and Alemán 2007; Padding and Briels 2011). However as most of these studies used a large degree of coarse graining ($\lambda \sim 20$) to study dynamics, an extra constraint is needed to prevent chains cutting through each other (Padding and Briels 2002). Here we focus on systems with fewer methylene groups per CG bead ($2 \leq \lambda \leq 6$) where we could largely (but not completely as discussed below) avoid including extra constraints to avoid chains cutting each other.

Figure 1 illustrates how the CG models with $\lambda = 2$–$6$ methylene groups per CG bead represent the underlying atomistic configuration. Though the chemical structure of PE is simple, it is a thermoplastic material useful in many applications, with tunable mechanical properties determined by the degree of branching. Using these CG models, we show here that one can capture polymer chain dynamics for long entangled polymers for time scales of order 1 ms using models that accurately represent atomistic detail. Accessing these large length and time scales, which are simply not accessible using fully atomistic models, allows one to measure a wide range of properties from the single-chain dynamics to the stress relaxation function and shear viscosity which depend on a hierarchy of length and time scales.



**Fig. 1** $C_{12}H_{24}$ segment of a PE chain represented with degree of coarse graining $\lambda = 2, 3, 4$, and 6 methylene groups per CG bead. The bead diameter corresponds to the position of the minimum in the nonbonded interaction for each CG model
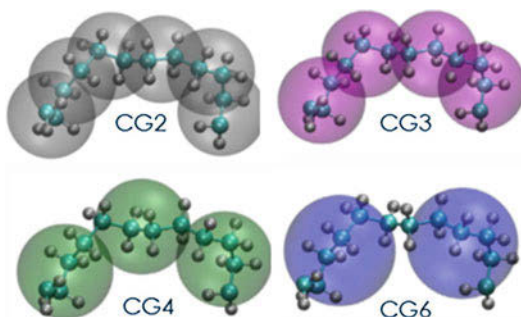
## 2        Model and Methodology

Each of the CG potentials was derived from a fully atomistic molecular dynamics simulation of a melt of $C_nH_{2n+2}$ with n = 96 for $\lambda$ = 2, 3, 4, and 6 and n = 95 for $\lambda$ = 5 (Salerno et al. 2016a, b; Peters et al. 2017). The atomistic simulations used the all-atom optimized potentials for liquid simulations (OPLS-AA) potential (Jorgensen et al. 1984, 1996) with modified dihedral coefficients (Siu et al. 2012). These modified OPLS-AA parameters reproduce the experimental static and dynamic chain properties for long alkanes better than the original OPLS-AA parameters. Tabulated CG bond potentials $U_B(l)$ and angle potentials $U_A(\theta)$ were determined by Boltzmann inversion of the atomistic bond length and angle distributions:

$$U_B(l) = -k_B T \log \left[ \frac{P(l)}{l^2} \right]$$

$$U_A(\theta) = -k_B T \log \left[ \frac{P(\theta)}{\sin \theta} \right],$$

where $l$ is the bond length for CG beads overlaid on the atomistic reference configurations and $\theta$ is the angle between CG bead triplets from the atomistic reference configuration. These two potentials were determined at temperature T = 400 and 500 K and found to be independent of temperature (Peters et al. 2017).In the current study, dihedral interactions were not included. However, recently Salerno and Bernstein (2018) have shown that for differences in the end-to-end distance, $<R^2>$ of up to 40% can occur for CG models with no dihedral interaction. This error is a result of correlations inherent to the CG models that can be represented as a dihedral interaction. Including a dihedral interaction in CG models of PE can effectively correct this error in the chain stiffness.

Tabulated nonbonded potentials were calculated by a multistep iterative Boltzmann inversion process (Müller-Plathe 2002; Reith et al. 2003; Voth 2008). The intermolecular radial distribution function g(r) from the atomistic simulation was used as the target for iteration of the nonbonded potentials. The resulting potentials, while giving excellent agreement between the CG g(r) and the target g(r), always resulted in a pressure which was significantly larger than that of the atomistic system. A pressure correction (Milano and Müller-Plathe 2005; Sun and Faller 2005; Wang et al. 2009) was then applied to bring the pressure of the CG and atomistic systems into agreement. This resulted in a slight increase of the pair correlation function in the vicinity of the first peak for the CG model compared to the target, but this difference was found to have no effect on the chain mobility and viscoelastic response of the system (Salerno et al. 2016a, b; Peters et al. 2017). The resulting potentials for four values of $\lambda$ at 500 K are shown in Fig. 2. Unfortunately, these CG potentials are not necessarily transferable to other temperatures. For example, with $\lambda$ = 4 the attractive well for the CG potential developed at 500 K is

**Fig. 2** Potentials for nonbonded interactions for $\lambda = 2$ (red circle), 3 (purple square), 4 (green triangle), and 6 (blue diamond)
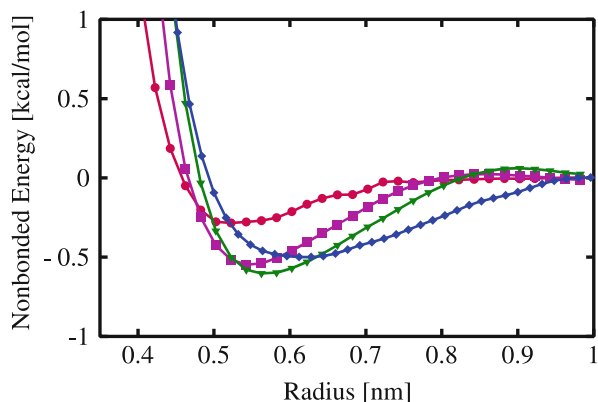
20% deeper compared with that developed at 400 K (Peters et al. 2017). Comparison of these potentials with the standard Lennard-Jones (LJ) 12:6 potential, which is often used in bead−spring models of polymers (Kremer and Grest 1990; Grest 2016), shows that these atomistically inspired CG potentials are softer with a much shallower attractive well than the LJ 12:6 potential (Peters et al. 2017). Since PE is locally stiff, as the degree of coarse graining increases, the methyl groups in each CG bead take up less of the volume. As a result, the nonbonded potentials become softer with increasing $\lambda$.

For large degrees of coarse graining, one must also include extra constraints so that the chains cannot cut through each other (Padding and Briels 2002). For polyethylene, even the $\lambda=6$ model has a surprisingly large equilibrium bond distance relative to the bead diameter which allows chains to occasionally cut through each other. Therefore, in our simulations, we added a modified soft segmental repulsive potential (Sirk et al. 2012) to CG beads to inhibit chain crossing for $\lambda = 6$. Complete details of the model and methodology are given in Salerno et al. (2016a, b).

One important advantage of CG models is that by eliminating the finest degrees of freedom, the time step in a molecular dynamics simulation is significantly larger than for fully atomistic models. We found that one can use a time step $\delta t = 20$ fs for $\lambda = 4$, 5, and 6 and 10 fs for $\lambda = 3$. However, for $\lambda = 2$, $\delta t$ is only 4 fs, comparable to that commonly used for atomistic (1 fs) and UA (1–2 fs) models. The reduction in the number of degrees of freedom in a system also creates a smoother free-energy landscape compared with fully atomistic or UA model simulations. While strong frictional and stochastic forces can be used to slow down the dynamics of the CG model to match those of the atomistic model (Salerno et al. 2016a, b), one can take advantage of this increase in the local dynamics of the CG model to simulate effectively much longer time scales (Harmandaris and Kremer 2009; Fritz et al. 2011).While the former approach may be useful for coarse graining small molecules, for entangled polymers which already have inherently slow dynamics, this additional speedup for CG models is very advantageous. Combining the

reduction in the number of atoms that one must simulate, the significant larger simulation time steps, and the increase in the dynamics from the smoother free-energy surface, the effective simulation time is effectively increased by at least three orders of magnitude for $\lambda = 4\text{--}6$, allowing one to reach times scales not accessible using fully atomistic simulations. Computationally for PE, $\lambda = 4$ and 5 are the most efficient since $\lambda = 6$ requires the addition of an extra bead between each CG bead to avoid chains cutting through each other. All of the simulations presented here were performed using the Large Atomic Molecular Massively Parallel Simulator (LAMMPS) molecular dynamics simulation code (Plimpton 1995), though any MD software package which allows tabulated forces for bond, angle, and nonbonded potentials could be used.

## 3    Results

In Salerno et al. (2016a, b), we probed the dynamics of polymers as the number of atoms included in a CG bead is varied from $\lambda = 2$ to 6. We found that independent of the degree of coarse graining, the static and dynamic properties are similar once the dynamic scaling factor $\alpha$ and non-crossing constraint for $\lambda = 6$ are included. Using these CG models, we have been able to reach times of order 1 ms, allowing us to measure several quantities which can be compared directly to experiments, including the stress relaxation function, plateau modulus, and shear viscosity. Here, some examples of our results are presented to illustrate the power of the coarse-graining methodology to sample times and length scales not accessible by atomistic simulations while retaining chemical specificity.

Coarse graining reduces the number of degrees of freedom in a system, creating a smoother free-energy landscape compared with fully atomistic simulations (Harmandaris and Kremer 2009; Fritz et al. 2011). This can be seen by measuring the mean squared displacement (MSD) for the atomistic model of PE with n = 96 and 480 carbons to the equivalent CG model as shown in Fig. 3a. The mobility of the chains in the CG models is clearly larger than in atomistic simulations. By scaling the time for each of the CG models, the results fall on a single collapsed curve for each chain length for the atomistic and CG models as shown in Fig. 3b. Notably, a single scaling factor $\alpha$ is required for each $\lambda$, independent of chain length. For T = 500 K, $\alpha$ varied from 6 to 9 for $\lambda = 2\text{--}6$ and increases with decreasing temperature (Peters et al. 2017). For $\lambda = 4$, $\alpha = 6.2$ for T = 500 K and increases to $\sim 12$ at 400 K. Combining the effects of fewer interaction sites, a larger time step, and the dynamic scaling factor $\alpha$, these CG models allow one to study well-entangled chains for long times.

Results for the MSD of the center of mass $g_3(t) = \langle (\mathbf{r}_{cm}(t) - \mathbf{r}_{cm}(0))^2 \rangle$ and motion of the center four beads $g_1(t) = \langle (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2 \rangle$ for chain lengths n = 1920, 2560 and 4000 for $\lambda = 4$ at 500 K are shown in Fig. 4. The experimental entanglement length for PE is 1.1–1.2 kg/mol or n $\sim$ 80 (Fetters et al. 1999; Vega et al. 2004). For n = 1920 and 2560, the MSD has reached the diffusive regime where MSD $\sim t^1$. Over intermediate time scales for all three systems, the chains
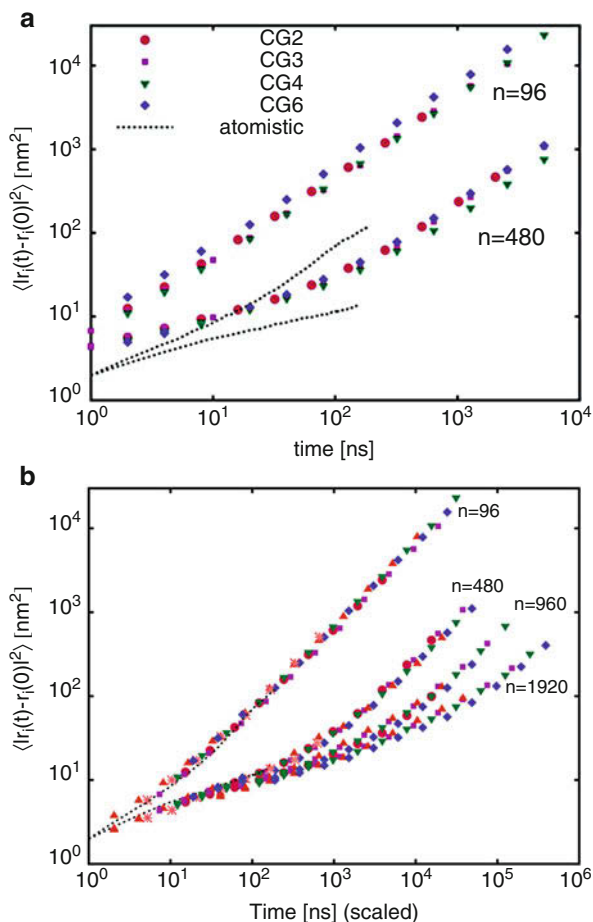
**Fig. 3** (**a**) Mean squared displacement of the inner 24 -CH$_2$- groups of each polymer chain at 500 K for different levels of coarse graining compared to atomistic model for n = 96 and 480. (**b**) Coarse-grained MSD data scaled by dynamic rescaling factor α for n = 96–1920 compared to atomistic simulations

show the expected t$^{1/4}$ scaling predicted the tube model (de Gennes 1971; Doi and Edwards 1986). The diffusion constant D = g$_3$(t)/6t for t > τ$_d$ is shown in Fig. 5 as a function of n. For large n, D follows a power law decay D = D$_1$n$^{-2.18}$ where D$_1$ = 3.08 × 10$^{-6}$ m$^2$/s. The decay of D with a power law greater than 2 for large M is consistent with experimental results (Lodge 1999).

From the crossover time $t_e^*$ from the early time t$^{1/2}$ Rouse regime to t$^{1/4}$ reptation, one can extract the tube diameter d$_T$ and the entanglement time $\tau_e$. Assuming the distribution of segment displacement along the tube is Gaussian on the scale of the tube diameter $d_T$, then the entanglement time $\tau_e = \frac{9}{\pi} t_e^*$ (Hou 2017). From Fig. 4, this crossover time is $t_e^* \sim 14$ ns, which gives $\tau_e \sim 40$ ns. Note that in the

**Fig. 4** Mean squared displacement of the center of mass $g_3(t)$ (closed) and center four CG beads $g_1(t)$ (open) for n = 1920 (black circles), 2560 (red squares), and 4000 (green triangles) for $\lambda = 4$. The dashed lines represent the scaling predictions $t^1$ for the diffusive regime and $t^{1/4}$ for the reptation regime
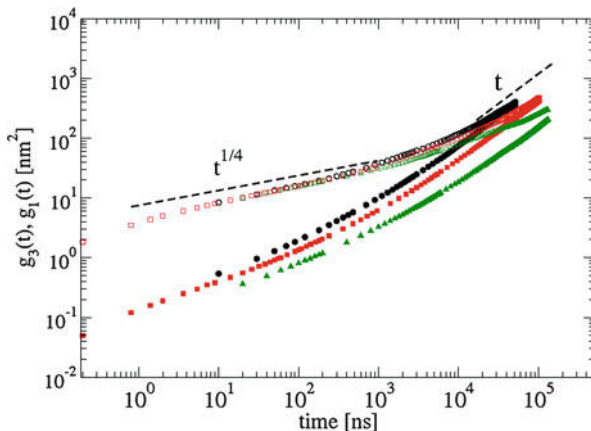


**Fig. 5** Diffusion constant D versus n for $\lambda = 4$ at 500 K
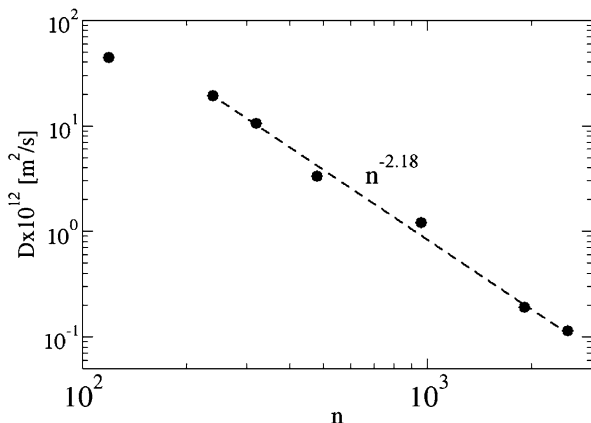


literature, $\tau_e$ and $t_e^*$ are assumed to be the same. The MSD of the center monomers at $t_e^*$, $g_{1e}^* = \frac{2}{3\pi}d_T^2$ gives a tube $d_T \sim 4.9$ nm (Hou 2017). If one neglects the $\frac{2}{3\pi}$ prefactor between $g_{1e}^*$ and $d_T^2$, which is often done in the literature (Hsu and Kremer 2016), then $d_T \sim 2.3$ nm. The latter value is in good agreement with the assumption that $d_T^2 = 2 R_g^2 (N_e)$, which gives $d_T \sim 2.2$ nm. Neutron spin-echo experiments by Richter et al. (1992) and Schleger et al. (1998) estimate $\tau_e \sim 5$ ns and tube diameter $d = \sqrt{3}d_T \cong 4.35$ nm or $d_T \sim 2.5$ nm (Hsu and Kremer 2017). Thus the crossover times and distances of the CG model capture the essential aspect of the polymer motion, demonstrating that one can capture long-time and length scales with CG models while accounting for atomistic detail.

The stress response function after a small perturbation $G(t)$ is one of the most important experimental measurements for polymers. For long entangled polymers, at short times $G(t)$ decays as the chains locally relax in response to the perturbation like any fluid. However, for intermediate times, $G(t)$ plateaus at $G_N^o = \frac{4}{5}\rho RT/M_e$ where $M_e$ is the entanglement molecular weight. This plateau region in $G(t)$

**Fig. 6** Stress autocorrelation function $G(t)$ for $\lambda = 4$ at 500 K for n = 1920 (black) and 4000 (green). Solid lines are fit to the Likhtman-McLeish formula (Likhtman and McLeish 2002)
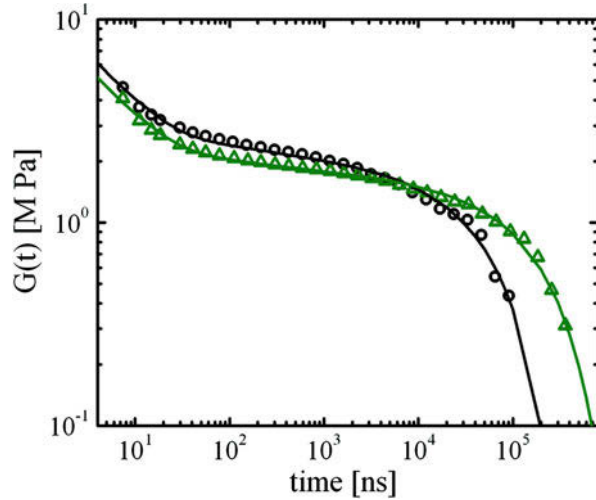
occurs for intermediate times where the chains are assumed to move in a tube due to entanglements from the other chains. Only after the chains have reached the diffusive regime, does $G(t)$ relax to zero. The relaxation modulus for each of our CG models was measured for different chain lengths via equilibrium stress autocorrelations. Figure 6 shows $G(t)$ for n = 1920 and 4000 for $\lambda = 4$, where time has been scaled by the dynamic rescaling factor $\alpha$. The solid lines in Fig. 4 are fit to the Likhtman-McLeish (LM) expression (Likhtman and McLeish 2002):

$$G(t) = \frac{k_B T}{nv} \left[ \frac{1}{5} \sum_{p=1}^{Z} \left( 4\mu(t)R(t) + e^{-tp^2/\tau_R} \right) + \sum_{p=Z+1}^{N} e^{-2tp^2/\tau_R} \right]$$

where

$$\mu(t) = \frac{8}{\pi^2} \sum_{q=1,odd}^{\infty} \frac{1}{q^2} \exp\left( -\frac{q^2 t}{\tau_d} \right)$$

is the Doi-Edwards reptation stress relaxation function (Doi and Edwards 1986) with terminal relaxation time $\tau_d$ and the double-reptation expression for constraint release $R(t) = \mu(t)$ is assumed (Marrucci 1985).The key quantity in this expression is the single-chain memory function $\mu(t)$ for the fraction of the primitive chain which has not escaped from its original tube after a time t. The fitting parameters are the number of entanglements per chain $Z$, the terminal relaxation time $\tau_d$, and the Rouse time $\tau_R$. The volume of a $CH_2$ monomer is $v = 0.031 nm^3$ for $\rho = 0.76 g/cm^3$. The best-fit results are $Z = 23 \pm 1$, $\tau_d = (1.49 \pm 0.04) \times 10^5 ns$, and $\tau_R = (2.60 \pm 0.16) \times 10^4 ns$ for n = 1920 and $Z = 40 \pm 1$, $\tau_d = (5.55 \pm 0.09) \times 10^5 ns$, and $\tau_R = (8.18 \pm 0.32) \times 10^4 ns$ for

$n = 4000$. Based on these results, the plateau modulus $G_e = 4Zk_BT/5nv = 2.3$MPa for $n = 1920$ and 1.8MPa for $n = 4000$; the entanglement time $\tau_e = \tau_R/Z^2 = 49.1$ ns for $n = 1920$ and 51.1 ns for $n = 4000$, consistent with the entanglement times extracted above from the MSD. The entanglement molecular weight $M_e = n/Z \times 14$g/mol $= 1.2$kg/mol for $n = 1920$ and 1.4kg/mol for $n = 4000$, which is consistent with experimental results for $M_e$ (Fetters et al. 1999; Vega et al. 2004). For comparison, the longest chain system we could study using fully atomistic simulations is n = 480, which is not long enough and cannot be simulated long enough to observe a plateau in *G(t)*.

Polymer entanglements can be directly determined by the primitive path analysis (PPA), an algorithm first developed by Everaers et al. (2004) In the PPA, the chain ends are fixed, and the intrachain excluded volume interactions are turned off while retaining the interchain excluded volume interactions. The energy of the system is then minimized by slowly cooling the system to T = 0. The entanglement molecular weight $M_e$ can then be obtained from the average contour length of the primitive path and the end-to-end distance of the chain (Everaers et al. 2004). In the standard PPA method (Everaers et al. 2004), the chains maintain the same diameter during the length minimization process. Alternatively one can introduce extra beads to reach the limit of zero thickness chains (Hoy and Grest 2007). These two methods are often referred to as thick- and thin-chain PPA, respectively. Although both methods can give good estimates of the entanglement length in a polymer melt, the effect of the bead size or CG level $\lambda$ should be considered. In this context, it is important to be able to represent a single reference configuration with various $\lambda$ and calculate the same primitive path length, independent of $\lambda$. We found that the thick-chain PPA method produces results that vary by 15–20% with $\lambda$. Hence, we adopt a procedure that is more like the thin-chain PPA method to reliably reproduce the primitive path length and entanglement length, independent of the coarse-graining level $\lambda$.

The primitive path contour length and the resulting entanglement length were estimated by inserting four (or nine) beads between each bead for $\lambda = 4$, setting the bead diameters to 1.3 Å (0.65 Å) and reducing the backbone bond length at constant tension. The primitive path contour length was computed for n = 480 for ten independent melt configurations. The primitive path length was estimated to be 19.8 nm, with lower and upper estimates of 24 and 31 CG beads for the entanglement length, based on the estimator proposed by Hoy et al. (2009). These thin-chain PPA values correspond to entanglement mass $M_e = 1.34$ and 1.74 kg/mol, somewhat higher than the estimate based on the plateau modulus.

Using non-equilibrium molecular dynamics simulations, we measured the shear viscosity $\eta$ as a function of shear rate over a wide range of shear rates from the shear-independent regime at low shear rates to the shear-thinning regime at high shear rates. Results for $\eta$ versus scaled shear rate for n = 96 to 1920 are presented in Salerno et al. (2016). Our results for the zero shear rate viscosity $\eta$ showed a crossover from a $n^1$ power law for small n to $n^{3.4}$ for large n at $n_c \sim 250$ or molecular weight $\sim 35$ kg/mol in excellent agreement with the experiment (Ferry 1980; Graessley and Edwards 1981).

While most studies of polymer melts model a homopolymer in which all the chains have the same length or at most a binary mixture of two lengths, experimental systems are never uniform. They are highly dispersed with a range of chain lengths. The standard way to characterize molar-mass dispersity is the dispersity index $Đ_M$ defined by $Đ_M = M_w/M_n$, where $M_w$ is the weight average molecular weight and $M_n$ is the number average molecular weight. $M_n$ is more sensitive to molecules of low molecular mass, while $M_w$ is more sensitive to molecules of high molecular mass. The best case experimentally for long polymers is $Đ_M \sim 1.02$ but in some cases $Đ_M > 2$. Even for $Đ_M \sim 1.02$, the ratio of the largest to shortest chain is larger than 2. How dispersity affects the chain mobility and viscoelastic response is difficult to predict theoretically, which makes testing theoretical models challenging due to the assumption that the system contains only uniform polymers.

Using our coarse-grained model with $\lambda = 4$, we (Peters et al. 2018) have recently initiated a study of the effect of polydispersity on polymer chain mobility and viscoelastic properties. We built dispersed systems of 2000 chains for $Đ_M = 1.02$, 1.04, and 1.08 and 4000 chains for $Đ_M = 1.16$ for highly entangled chains of molecular weight $M_w = 36$ kg/mol (n = 2650 $CH_2$ monomers or 640 CG beads). This chain length was chosen since it is long enough for the chains to be well entangled but short enough that the system fully relaxes on the time scales accessible to simulations ($\sim$600–800 μs). For short times, the monomers in the center of the chain diffuse the same independent of the degree of dispersity. However, for longer times, the MSD of the more dispersed samples moves faster. Unlike experiment, simulations can be used to extract more detailed information of how the various chain populations diffuse. For example, the average diffusion constant of all the chains as well as the shortest and longest can be readily determined as shown in Fig. 7. As $Đ_M$ increases, the spread in the diffusion constant increases very rapidly with $Đ_M$, while the average value of D increases by only a factor of 50% over this range.



**Fig. 7** Diffusion constant D as a function of molar-mass dispersity $Đ_M$ averaged over all the chains (black), for the shortest 5% of the chains (red), and longest 5% of the chains (blue) for n = 2560 for $\lambda = 4.5$
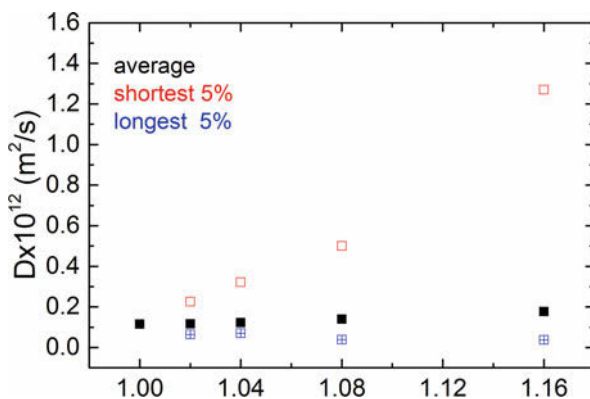
# 4    Conclusions

Here we probed the structure and dynamics of polymers with various numbers of atoms included in a coarse-grained bead. We have shown that for long, entangled polymers, independent of the degree of coarse graining, all static and dynamic properties are essentially the same once the dynamic scaling factor α and non-crossing constraint for λ = 6 are included. Due to the larger time step and reduced number of degrees of freedom of the coarse-grained model as well as the smoother free-energy surface, the coarse-grained simulations are effectively 3–4 orders of magnitude faster than the fully atomistic model, depending on the degree of coarse graining and the temperature. Of the five models studied, λ = 4 and 5 offer the optimum performance since larger degrees of coarse graining require an additional non-crossing constraint which negates the gain in effective speed due to free pairwise interactions. This translates to simulation times of hundreds to thousands of μs, which allows us to measure many quantities which can be compared directly to experiments, including the stress relaxation function, plateau modulus, and shear viscosity. With this speedup, one can readily extend the present study to longer chains as well as short- and long-chain branching in addition to the study of polydispersed systems.

# References

Abrams CF, Kremer K (2003) Combined coarse-grained and atomistic simulation of liquid bisphenol a−polycarbonate: liquid packing and intramolecular structure. Macromolecules 36:260–267

Ashbaugh HS, Patel HA, Kumar SK, Garde S (2005) Mesoscale model of polymer melt structure: self-consistent mapping of molecular correlations to coarse-grained potentials. J Chem Phys 122:104908

Chen L-J, Qian H-J, Lu Z-Y, Li Z-S, Sun C-C (2006) An automatic coarse-graining and fine-graining simulation method: application on polyethylene. J Phys Chem B 110:24093–24100

Curcó D, Alemán C (2007) Coarse-grained simulations of amorphous and melted polyethylene. Chem Phys Lett 436:189–193

Dallavalle M, van der Vegt NFA (2017) Evaluation of mapping schemes for systematic coarse graining of higher alkanes. Phys Chem Chem Phys 19:23034–23042

Doi M, Edwards SF (1986) The theory of polymer dynamics. Oxford University, Oxford, UK

Everaers R, Sukumaran SK, Grest GS, Svaneborg C, Sivasubramanian A, Kremer K (2004) Rheology and microscopic topology of entangled polymeric liquids. Science 303:823–826

Ferry JD (1980) Viscoelastic properties of polymers. Wiley, New York

Fetters LJ, Lohse DJ, Milner ST, Graessley WW (1999) Packing length influence in linear polymer melts on the entanglement, critical, and reptation molecular weights. Macromolecules 32: 6847–6851

Fritz D, Koschke K, Harmandaris VA, van der Vegt NF, Kremer K (2011) Multiscale modeling of soft matter: scaling of dynamics. Phys Chem Chem Phys 13:10412–10420

Fukunaga H, Takimoto J-i, Doi M (2002) A coarse-graining procedure for flexible polymer chains with bonded and nonbonded interactions. J Chem Phys 116:8183–8190

de Gennes P-G (1971) Reptation of a polymer chain in the presence of fixed obstacles. J Chem Phys 55:572–579

Graessley WW, Edwards SF (1981) Entanglement interactions in polymers and the chain contour concentration. Polymer 22:1329–1334

Grest GS (2016) Communication: polymer entanglement dynamics: role of attractive interactions. J Chem Phys 145:141101

Guerrault X, Rousseau B, Farago J (2004) Dissipative particle dynamics simulations of polymer melts. I. Building potential of mean force for polyethylene and cis-polybutadiene. J Chem Phys 121:6538–6546

Harmandaris VA, Kremer K (2009) Dynamics of polystyrene melts through hierarchical multiscale simulations. Macromolecules 42:791–802

Harmandaris VA, Reith D, van der Vegt NFA, Kremer K (2007) Comparison between coarse-graining models for polymer systems: two mapping schemes for polystyrene. Macromol Chem Phys 208:2109–2120

Hou J-X (2017) Note: determine entanglement length through monomer mean-square displacement. J Chem Phys 146:026101

Hoy RS, Grest GS (2007) Entanglements of an end-grafted polymer brush in a polymeric matrix. Macromolecules 40:8389–8395

Hoy RS, Foteinopoulou K, Kröger M (2009) Topological analysis of polymeric melts: chain-length effects and fast-converging estimators for entanglement length. Phys Rev E 80:031803

Hsu H-P, Kremer K (2016) Static and dynamic properties of large polymer melts in equilibrium. J Chem Phys 144:154907

Hsu H-P, Kremer K (2017) Detailed analysis of Rouse mode and dynamic scattering function of highly entangled polymer melts in equilibrium. Euro Phys J Special Topics 226:693–703

Jorgensen WL, Madura JD, Swenson CJ (1984) Optimized intermolecular potential functions for liquid hydrocarbons. J Am Chem Soc 106:6638–6646

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236

Karimi-Varzabeh HA, van der Vegt NFA, Mueller-Plathe F, Carbone P (2012) How good are coarse-grained models? A comparison for actatic polystyrene. ChemPhysChem 13:3428–3439

Kremer K, Grest GS (1990) Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. J Chem Phys 92:5057–5086

Li Y, Abberton BC, Kröger M, Liu WK (2013) Challenges in multiscale modeling of polymer dynamics. Polymer 5:751–832

Likhtman AE, McLeish TC (2002) Quantitative theory for linear dynamics of linear entangled polymers. Macromolecules 35:6332–6343

Lodge TP (1999) Reconciliation of the molecular weight dependence of diffusion and viscosity in entangled polymers. Phys Rev Lett 83:3218

Marrucci G (1985) Relaxation by reptation and tube enlargement: a model for polydisperse polymers. J Polym Sci B Polym Phys 23:159–177

Martin MG, Siepmann JI (1998) Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. J Phys Chem B 102:2569–2577

Maurel G, Schnell B, Goujon F, Couty M, Malfreyt P (2012) Multiscale modeling approach toward the prediction of viscoelastic properties of polymers. J Chem Theory Comput 8:4570–4579

Maurel G, Goujon F, Schnell B, Malfreyt P (2015) Prediction of structural and thermomechanical properties of polymers from multiscale simulations. RSC Adv 5(19):14065–14073

Milano G, Müller-Plathe F (2005) Mapping atomistic simulations to mesoscopic models: a systematic coarse-graining procedure for vinyl polymer chains. J Phys Chem B 109:18609–18619

Müller-Plathe F (2002) Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back. ChemPhysChem 3:754–769

Nath SK, Escobedo FA, de Pablo JJ (1998) On the simulation of vapor–liquid equilibria for alkanes. J Chem Phys 108:9905–9911

Padding J, Briels WJ (2001) Uncrossability constraints in mesoscopic polymer melt simulations: non-Rouse behavior of $C_{120}H_{240}$. J Chem Phys 115:2846–2859

Padding J, Briels WJ (2002) Time and length scales of polymer melts studied by coarse-grained molecular dynamics simulations. J Chem Phys 117:925–943

Padding J, Briels WJ (2011) Systematic coarse-graining of the dynamics of entangled polymer melts: the road from chemistry to rheology. J Phys Condens Matter 23:233101

Paul W, Yoon DY, Smith GD (1995) An optimized united atom model for simulations of polymethylene melts. J Chem Phys 103:1702–1709

Peter C, Kremer K (2009) Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back. Soft Matter 5:4357–4366

Peters BL, Salerno KM, Agrawal A, Perahia D, Grest GS (2017) Coarse grained modeling of polyethylene melts: effect on dynamics. J Chem Theory Comp 13:2890–2896

Peters BL, Salerno KM, Ge T, Perahia D, Grest GS (2018) Dynamics of polydispersed, entangled polymer melts. in preparation

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117:1–19

Reith D, Pütz M, Müller-Plathe F (2003) Deriving effective mesoscale potentials from atomistic simulations. J Comput Chem 24:1624–1636

Richter D, Butera R, Fetters L, Huang J, Farago B, Ewen B (1992) Entanglement constraints in polymer melts. A neutron spin echo study. Macromolecules 25:6156–6164

Salerno KM, Bernstein N (2018) Persistence length, end-end distance and structure of coarse-grained polymers. J Chem Theory Comp 14:2219–2229

Salerno KM, Agrawal A, Perahia D, Grest GS (2016a) Resolving dynamic properties of polymers through coarse-grained computational studies. Phys Rev Lett 116:058302

Salerno KM, Agrawal A, Peters BL, Perahia D, Grest GS (2016b) Dynamics in entangled polyethylene melts. Euro Phys J Special Topics 225:1707–1722

Schleger P, Farago B, Lartigue C, Kollmar A, Richter D (1998) Clear evidence of reptation in polyethylene from neutron spin-echo spectroscopy. Phys Rev Lett 81:124–127

Siepmann JI, Karaborni S, Smit B (1993) Simulating the critical properties of complex fluids. Nature 365:330–332

Sirk TW, Slizoberg YR, Brennan JK, Lisal M, Andzelm JW (2012) An enhanced entangled polymer model for dissipative particle dynamics. J Chem Phys 136:134903

Siu SW, Pluhackova K, Böckmann RA (2012) Optimization of the OPLS-AA force field for long hydrocarbons. J Chem Theory Comp 8:1459–1470

Sun Q, Faller R (2005) Systematic coarse-graining of atomistic models for simulation of polymeric systems. Comput Chem Eng 29:2380–2385

Vega JF, Rastogi S, Peters GWM, Meijer HEH (2004) Rheology and reptation of linear polymers. Ultrahigh molecular weight chain dynamics in the melt. J Rheol 48:663–678

Voth GA (2008) Coarse-graining of condensed phase and biomolecular systems. CRC press, Boca Raton

Wang H, Junghans C, Kremer K (2009) Comparative atomistic and coarse-grained study of water: what do we lose by coarse-graining? Euro Phys J E 28:221–229

# Top-Down Hybrid Models of Polymers

# 64

Kostas Ch. Daoulas

## Contents

**Abstract**

We review basic conceptual and methodological aspects of a special class
of hybrid models of polymeric materials. These particle-based models are
mesoscopic descriptions, where coarse-grained "force fields" are introduced
through functionals of local order parameters. These functionals bear similarities
to free-energy models in classical density functional theory and phase-field
simulations. Introducing interactions through free-energy-like functionals allows
one to develop top-down particle-based models with desired thermodynamic
behavior in a systematic and modular way. Hybrid models serve as a framework
for efficient simulations based on standard tools such as Monte Carlo sampling
and dynamical algorithms. Thanks to their special structure, hybrid models also
provide a basis for field-theoretical methods, such as numerical self-consistent
field theory.

K. C. Daoulas (✉)
Max Planck Institute for Polymer Research, Ackermannweg 10, Mainz, Germany
e-mail: daoulas@mpip-mainz.mpg.de

# 1    Introduction

Investigating the mesoscopic behavior of polymeric materials with computer simulations requires special techniques. Considering that the involved length scales are at least on the order of 100 nm, these techniques must handle samples containing thousands of long polymer chains, while circumventing the slow polymer dynamics. Both requirements can be satisfied using drastically coarse-grained (CG) models, where large groups of building units of the actual polymer chains are represented by a single effective particle. Projecting a large number of microscopic states on a single CG configuration leads to strong entropic contributions to potentials acting between drastically CG particles. Therefore effective interactions have an explicit free-energy character, and their strength is comparable to the thermal energy, $kT$. Soft interactions and reduced amount of degrees of freedom allow for an efficient sampling of the configuration space.

The strong thermodynamic nature of interactions in drastically CG models justifies the implementation of computationally efficient soft potentials but complicates their determination. Systematic bottom-up methods, constructing the potentials such that the CG model reproduces physical quantities known from a more detailed description of the system (Kremer and Müller-Plathe 2002; Noid et al. 2008), are of limited use. Some of the reasons lie in representability and transferability problems (Peter and Kremer 2010; Noid et al. 2008) which are amplified by the prominent thermodynamic nature of drastically CG potentials. Tiny inaccuracies in bottom-up determination of interactions on the level of a single CG repeat unit change significantly phase behavior – due to chain connectivity, the impact of such inaccuracies increases proportionally to the degree of polymerization.

Top-down development of drastically CG models presents an alternative to bottom-up coarse-graining. Top-down approaches construct the CG model *postulatively*, building into the model "by hand" features that are necessary to reproduce a set of known structural, dynamic, and/or thermodynamic properties. The choice of postulated properties depends on the addressed questions, insights from fundamental physical principles, and/or experimental data. Top-down models first serve as a tool for qualitative understanding and prediction of unknown properties, based on existing limited knowledge about a system. Nevertheless, for certain classes of materials, top-down drastically CG models can provide quantitative predictions (Lodge 2003).

Comparing to bottom-up strategies, top-down coarse-graining is less formalized. Intriguingly, various structural, dynamic, and thermodynamic properties in polymers are frequently governed by universal laws (de Gennes 1979). These laws account for specific material chemistry through a few characteristic parameters, known as "invariants." Universality plays a central role in top-down modeling, enabling the study of polymer materials through very simple (minimal Müller et al. 2006) models, which reproduce the few required invariants. However, for many materials (especially those characterized by complex, heterogeneous structure),

finding universal laws is not straightforward. Moreover, universality may not be applicable altogether because the mesoscopic behavior remains coupled (possibly weakly) to microscopic features of molecular architecture and interactions. Therefore, it is necessary to develop systematic methods for constructing top-down models which are more complex compared to those based on simple universal laws. It is desirable that these top-down approaches are modular, allowing one to add to the model additional features to improve accuracy or expand the applicability range.

In this chapter, we will review the basics of top-down models developed through a systematic and modular strategy which is gaining popularity over the last years. The method constructs particle-based CG models in a hybrid manner, by introducing nonbonded interactions through a functional of a few local order parameters (collective coordinates). The functional is structurally similar to free energies in classical density functional theory (DFT) and phase-field models but differs conceptually, being a functional of *instantaneous* values of collective coordinates. Connecting instantaneous order parameters to individual coordinates of CG particles allows one to transform the functional into a law for nonbonded energies or forces and employ standard simulation techniques. The strength of the method is that the thermodynamic behavior of the model is modulated on the level of the functional used to "template" the nonbonded interactions. The choice of an appropriate functional is facilitated by a number of theoretical techniques including classical DFT and self-consistent field (SCF) theory.

## 1.1    Ingredients of Hybrid Top-Down Models

The first step of hybrid top-down coarse-graining is to specify a scheme for replacing groups of microscopic degrees of freedom by CG particles. The choice of the scheme depends on the features to be included into the model. The interactions between the CG particles are separated into bonded and nonbonded, defined by effective Hamiltonians $H_b$ and $H_{nb}$. Because the molecular structure is explicitly described on particle-based level, $H_b$ are functions of particle coordinates. Depending on the mapping scheme, $H_b$ is defined using either generic functions or chemistry-specific potentials, which can be derived, e.g. from standard bottom-up coarse-graining of single-chain atomistic configurations. By splitting the CG interactions into bonded and nonbonded contributions, hybrid top-down models follow the usual ansatz of coarse-graining (Peter and Kremer 2010) which assumes that intramolecular CG potentials do not depend on the environment of the molecules. Frequently, this assumption does not compromise the usefulness of top-down CG models for qualitative predictions. However, a posteriori refinement of parameters entering $H_b$ is usually necessary for quantitative studies.

After the mapping scheme and the bonded interactions are defined, $H_{nb}$ is introduced using a functional of collective coordinates. The concept is best clarified considering the formal statistical mechanics of hybrid models, where configurational partition functions take the form:

$$Z(n, V, T) =$$

$$\frac{1}{n!} \int d\mathbf{R} \exp \Big[ - \sum\nolimits_{i=1}^{n} \beta H_b(\mathbf{R}_i) - \underbrace{\int d\mathbf{r} f(\{\hat{\rho}_\lambda(\mathbf{r}; \mathbf{R}_1, \cdots, \mathbf{R}_n)\})}_{\beta H_{nb}} \Big] \quad (1)$$

Here, without loss of generality, we consider the canonical ensemble and assume that the polymer system contains $n$ identical chains (extending to several chain species, e.g., homopolymer blends, is straightforward). Nevertheless, the identical polymers can contain several monomer species. The coordinates of CG monomers comprising the $i$-th chain are denoted by $\mathbf{R}_i$, and $\int d\mathbf{R} \equiv \int \prod_{i=1}^{n} d\mathbf{R}_i$ indicates integration over all possible realizations of monomer coordinates. The argument of $H_b(\mathbf{R}_i)$ emphasizes the explicit particle-based description of molecular architecture. Expressing $H_{nb}$ through a functional of local order parameters makes Eq. 1 to appear different from the partition functions of standard CG models. For illustration purposes, we consider the simplest case, where the functional depends only on the set of local densities of the different monomer species, $\{\hat{\rho}_\lambda(\mathbf{r}; \mathbf{R}_1, \cdots, \mathbf{R}_n)\}$ (the index $\lambda$ spans the spectrum of species present in the system). The "hats" over the densities indicate that we are dealing with their *instantaneous* values and that they are not the fundamental degrees of freedom. They are operators transforming a configuration of coordinates $\mathbf{R}$ into instantaneous values of collective coordinates at point $\mathbf{r}$. To emphasize the role of $\hat{\rho}_\lambda$ as operators, we write explicitly their arguments $\mathbf{R}_1, \cdots, \mathbf{R}_n$. In the following, we will omit these arguments and use the compact notation $\hat{\rho}_\lambda(\mathbf{r})$. Through the operators, $H_{nb}$ depends implicitly on coordinates. Therefore (Edwards 1965; Fredrickson 2006) Eq. 1 still defines a particle-based partition function and is not a field-theoretical description.

Simulations based on models formalized through Eq. 1 require energies or forces which are functions of monomer coordinates $\mathbf{R}$. Therefore, the mathematical form of the functional and operators must facilitate the derivation of energies or forces from $H_{nb}$. The mathematical structure of these two ingredients determines the simulation techniques that can be applied to the given model.

## 1.2    Defining Functionals

An attractive idea for constructing $H_{nb}$ is to benefit from the classical DFT where free-energy functionals are playing a central role. Classical DFT is founded upon a theorem stating that for a system of particles (atoms or molecules) with defined interactions, the part of the Helmholtz free energy which does not explicitly depend on external fields (possibly acting on the system) is a unique functional of equilibrium local average densities, $\{\bar{\rho}_\lambda(\mathbf{r})\}$. This free-energy component, $F_{int}(\{\bar{\rho}_\lambda(\mathbf{r})\})$, is known as the "intrinsic Helmholtz free energy" and constitutes an inherent property of the system. The total free energy is obtained by adding to the intrinsic part trivial terms accounting for interactions with external fields. An analog of this theorem was first derived in quantum mechanics considering the density of

an interacting gas of electrons at ground state (Hohenberg and Kohn 1964) and thermal equilibrium ($k_B T \neq 0$) (Mermin 1965). For classical systems, detailed proofs are available in specialized reviews and textbooks; see Evans (1979), Hansen and McDonald (2006), and Plischke and Bergersen (1994). Although the functional of the intrinsic Helmholtz free energy formally exists, establishing this functional from a microscopic description is a formidable task. Developing strategies for obtaining approximate free-energy functionals is an active area of research in modern DFT (Wu and Li 2007; Löwen 2002).

The intrinsic Helmholtz free energy from DFT can be used to define nonbonded interactions in a hybrid model by setting $H_{nb}$ equal to $F_{int}(\hat{\rho}_\lambda(\mathbf{r}))$. However, there are important subtleties. First when $F_{int}(\{\hat{\rho}_\lambda(\mathbf{r})\})$ is used to define $H_{nb}$, density-dependent terms associated with translational and conformational entropy must be excluded because particle coordinates and molecular structure are explicitly present in the model. Otherwise, entropic contributions will be double counted. Second the intrinsic free energy of a system described by a model where $H_{nb}$ is determined by $F_{int}(\{\hat{\rho}_\lambda(\mathbf{r})\})$ differs from the original DFT free energy $F_{int}(\{\bar{\rho}_\lambda(\mathbf{r})\})$ (even when redundant entropic terms are omitted). The discrepancy stems from fluctuations captured by the partition function, Eq. 1. One consequence of the difference of the two free energies is that the top-down model reproduces the thermodynamic behavior corresponding to the DFT free energy only approximately.

To clarify better the origin of differences in the free energies, it is instructive to consider a system with $n$ interacting identical monomeric particles. Even in this simple example, the average local equilibrium density, $\bar{\rho}(\mathbf{r})$, can be inhomogeneous; coexisting liquid and vapor phases are an example. Assuming that for this system the DFT free energy is known, we define a hybrid toy model with a partition function:

$$Z(n, V, T) = \frac{1}{\Lambda^{3n} n!} \int d\mathbf{R}_1 \cdots d\mathbf{R}_n \exp\left[ -\beta F_{int(ex)}(\hat{\rho}(\mathbf{r})) \right] \tag{2}$$

$$= \frac{1}{\Lambda^3 n!} \int \mathscr{D}\rho(\mathbf{r}) \exp\left[ -\beta F_{int(ex)}(\rho(\mathbf{r})) \right] \int d\mathbf{R}_1 \cdots d\mathbf{R}_n \delta(\rho(\mathbf{r}) - \hat{\rho}(\mathbf{r})) \tag{3}$$

For clarity, we include the kinetic part of the partition function ($\Lambda$ is the thermal de Broglie wavelength). The definition of $H_{nb}$ is based on the excess DFT functional $\beta F_{int(ex)}(\hat{\rho}(\mathbf{r}))$, that is, the intrinsic free energy without the translational entropy term $\int d\mathbf{r}\hat{\rho}(\mathbf{r}) \left[ \ln(\hat{\rho}(\mathbf{r})\Lambda^3) - 1 \right]$. Equation 2 is transformed into an equivalent form by introducing an auxiliary integration in the functional space of all possible density distributions, as demonstrated in Eq. 3. The integration is denoted by $\int \mathscr{D}\rho(\mathbf{r})$ where $\delta(\rho(\mathbf{r}) - \hat{\rho}(\mathbf{r}))$ is a $\delta$-functional constraining the density field $\hat{\rho}(\mathbf{r})$. Because this density field is an implicit function of coordinates, the integral $\int d\mathbf{R}_1 \cdots d\mathbf{R}_n$, taken over the coordinates of the $n$ particles, quantifies the number of microstates with the same density pattern, $\rho(\mathbf{r})$. Hence, this conventional integration restores the translational entropy which was omitted from $F_{int(ex)}$. From Eq. 3, we obtain the free energy of our toy model as:

$$\beta F(n, V, T)$$

$$= -\ln\left(\int \mathcal{D}\rho(\mathbf{r})\exp\left[-\beta F_{\text{int(ex)}}(\rho(\mathbf{r}))\right]\underbrace{\int\frac{d\mathbf{R}_1\cdots d\mathbf{R}_n}{\Lambda^{3n}n!}\delta(\rho(\mathbf{r})-\hat{\rho}(\mathbf{r}))}_{=\exp(-\int d\mathbf{r}\rho(\mathbf{r})[\ln(\rho(\mathbf{r})\Lambda^3)-1])}\right)$$

$$(4)$$

Considering that the set of possible density distributions in $\int \mathcal{D}\rho(\mathbf{r})$ contains the pattern of the average density $\bar{\rho}(\mathbf{r})$, we conclude that:

$$\beta F(n, V, T) < \underbrace{\beta F_{\text{int(ex)}}(\bar{\rho}(\mathbf{r})) + \int d\mathbf{r}\bar{\rho}(\mathbf{r})\left[\ln\left(\bar{\rho}(\mathbf{r})\Lambda^3\right)-1\right]}_{\beta F_{\text{int}}(\bar{\rho}(\mathbf{r}))} \qquad (5)$$

The free energy of the toy model equals the DFT free energy only in the case of quenched fluctuations, where the instantaneous and the average density distributions are equal.

Frequently it is simpler to define $H_{\text{nb}}$ phenomenologically than to extract it through formal DFT derivations. For this purpose, one can benefit from generic relationships between the mathematical structure of the functional and thermodynamics. For example, defining $H_{\text{nb}}$ through a third-order density polynom is the simplest way to obtain models describing liquid-vapor coexistence (double-well free-energy landscape). This feature is important for systems where free polymer surfaces play an important role. Consequently, third-order functionals have been frequently applied for generic modeling of polymer brushes (Soga et al. 1995), films (Müller and Smith 2005) of melts and solutions, and amphiphilic membranes (Daoulas and Müller 2010). Although for some actual materials third-order functionals can approximately reproduce phase-equilibria properties (Binder et al. 2005), usually more complicated expressions are required. A simple empirical approach is to use higher-order polynomials, which are reminiscent of the Wohl's expansion (Wohl 1946; Prausnitz et al. 1999) of the excess Gibbs free energy.

Currently, most hybrid models for multicomponent polymer liquids employ $H_{\text{nb}}$ which are variants of the general form:

$$\beta H_{\text{nb}} = \frac{\kappa\rho_0}{2}\int d\mathbf{r}\,\left(\hat{\phi}_A(\mathbf{r}) + \hat{\phi}_B(\mathbf{r}) - 1\right)^2 + \chi\rho_0\int d\mathbf{r}\,\hat{\phi}_A(\mathbf{r})\hat{\phi}_B(\mathbf{r}) \qquad (6)$$

For simplicity, Eq. 6 refers to a system with only two types of monomers, A and B (extending to multiple components is straightforward). The operators $\hat{\phi}_\lambda(\mathbf{r}) = \hat{\rho}_\lambda(\mathbf{r})/\rho_0$ define the local volume fractions of the two monomer species ($\lambda = $ A,B). The average reference density is given by $\rho_0$.

The two-term structure of Eq. 6 illustrates the modularity of hybrid models. When A and B monomers are identical, the first term of $\beta H_{\text{nb}}$ describes a homopolymer melt, where a simple equation-of-state (EOS) penalizes deviations of

the instantaneous density from the reference value. The simple quadratic form can be motivated by a second-order expansion of the excess free energy of a positionally inhomogeneous homopolymer melt (Helfand 1975; Jiang et al. 2017) around a homogeneous reference state with density $\rho_0$. More sophisticated EOS can be integrated into the model replacing the first term by more complex functionals. The compressibility of the CG polymer liquid is set by the parameter $\kappa$. The second term of $\beta H_{nb}$ describes the incompatibility of the two components. The strength of the incompatibility is controlled by $\chi$. This standard description can be traced back to the lattice-based Flory-Huggins (FH) model, where the FH parameter $\chi$ is expressed through contact energies of monomers occupying neighboring lattice points. This expression can be also motivated assuming short-ranged pairwise interactions of monomers in continuum space (Doi and Edwards 1986; Matsen 2006; Müller 2006). From the standpoint of top-down modeling, the second term can be introduced as a simple penalty of contacts between unlike monomers, without considering any specific microscopic picture. The question of how one can determine $\chi$ to represent a given multicomponent polymer system has attracted significant attention (Wang 2002; Schweizer and Singh 1995; Müller 2006; Schmid 2011; Müller and Binder 1995; Morse and Chung 2009) and cannot be covered in this chapter. We mention briefly that determining $\chi$ from the "first principles" is a challenging task – being a coarse-grained interaction parameter, $\chi$ is in fact a free energy which depends on numerous subtle enthalpic and entropic effects present on the microscale. Within the spirit of top-down modeling, one usually follows the pragmatic approach where $\chi$ is considered as a phenomenological (black box) parameter, which is adjusted until selected mesoscopic observables reproduce their counterparts in experiments or simulations with more detailed models. Typical examples of these observables are intrinsic interfacial width, composition of coexisting phases, and structure factors.

It is instructive to compare $\beta H_{nb}$ from Eq. 6 with the standard Flory-Huggins-de-Gennes (de Gennes 1980) (FHG) density functional describing homopolymer blends. In the simplest case of a symmetric binary blend of $A$- and $B$-homopolymers, this functional takes the form:

$$\beta F(\{\phi_A\}) = \int d\mathbf{r} f(\phi_A(\mathbf{r})), \quad f(\phi_A(\mathbf{r})) = \sqrt{\bar{N}} \Big[ \phi_A(\mathbf{r}) \ln \phi_A(\mathbf{r})(1 - \phi_A(\mathbf{r}))$$

$$+ \ln(1 - \phi_A(\mathbf{r})) + \chi N \phi_A(\mathbf{r})(1 - \phi_A(\mathbf{r})) + \tilde{B}(\phi_A(\mathbf{r}))|\nabla \phi_A(\mathbf{r})|^2 \Big] \tag{7}$$

Eq. 7 assumes that the blend is incompressible with constant density $\rho_0$ so that the free-energy density, $f(\phi_A(\mathbf{r}))$, depends only on the concentration of one component, $\phi_A(\mathbf{r})$. All length scales are expressed in units of the root-mean-square end-to-end distance, $R_e$, of the homopolymer chains. The invariant degree of polymerization, $\bar{N}$, is defined such that $\sqrt{\bar{N}} \equiv n R_e^3 / V$, where $n/V$ is the number density of homopolymer chains. $\sqrt{\bar{N}}$ quantifies the amount of interacting chains threading through the volume of a test polymer coil and therefore sets the strength

of fluctuations. Comparing Eq. 6 with the FHG functional demonstrates that $\beta H_{nb}$ is missing the logarithmic terms which are associated with the translational entropy of the molecules. As discussed previously, omitting the logarithmic terms is consistent with the explicit presence of particle coordinates in the hybrid model. The nonlocal square-gradient term in the FHG functional penalizes the formation of interfaces and is, therefore, important for reproducing correct physics. Specifically the term accounts for the loss of conformational entropy (Lifshitz et al. 1978) of the chains at an $A/B$ interface. The precise form of $\tilde{B}(\phi_A)$ depends on the strength of segregation and is not essential for our discussion. Although the square-gradient term is missing from $\beta H_{nb}$, the hybrid model still captures the loss of conformational entropy at interfaces because the molecular architecture is explicitly described. However, introducing into $H_{nb}$ non-locality of interactions is important for modeling systems containing monomeric particles. Dilute films of polymer solutions are an example, where non-locality is required to describe a phase of liquid solvent exposed to vapor.

Simple symmetry arguments are useful when developing functional-based models for materials with complex order, such as polymer liquid crystals (LC). For uniaxial nematics, one can start from the Landau-de-Gennes expansion (de Gennes and Prost 1995) of the free energy with respect to the local alignment tensor, $\hat{\mathbf{Q}}(\mathbf{r})$:

$$\beta F = \int d\mathbf{r} \left[ \frac{A}{2} \hat{\mathbf{Q}}(\mathbf{r}){:}\hat{\mathbf{Q}}(\mathbf{r}) + \frac{B}{3} (\hat{\mathbf{Q}}(\mathbf{r}){\cdot}\hat{\mathbf{Q}}(\mathbf{r})){:}\hat{\mathbf{Q}}(\mathbf{r}) + \frac{C}{4} (\hat{\mathbf{Q}}(\mathbf{r}){:}\hat{\mathbf{Q}}(\mathbf{r}))^2 \right] \qquad (8)$$

The expansion in Eq. 2 includes terms justified by general symmetry arguments (de Gennes and Prost 1995) (rotational invariance of free energy). The constants $A$, $B$, and $C$ are unimportant for our discussion. Expressing the tensorial operators $\hat{\mathbf{Q}}(\mathbf{r})$ through coordinates and orientation vectors of CG units is straightforward. Therefore Eq. 8 can be used to generate particle-based LC models of polymer LC (Daoulas et al. 2012; Gemünden et al. 2013; Gemünden and Daoulas 2015). Retaining only the quadratic term of Eq. 8 delivers the class of Maier-Saupe models. The Landau-de-Gennes expansion can be generalized (Matteis et al. 2008) to include additional tensorial order parameters describing, e.g. biaxial order (Gemünden et al. 2013). We notice that the functional $\beta F$ is local without scalar combinations of gradients of $\hat{\mathbf{Q}}(\mathbf{r})$. These terms are necessary in a continuum description because they penalize inhomogeneities in ordering and are connected to Frank-Oseen elastic constants. However, Eq. 8 is sufficient for constructing particle-based models of polymer LC because non-locality is generated by explicit chain connectivity and using density distributions, which deliver potentials with finite range (see Sect. 1.5).

Continuum functional-based descriptions can address crystallization phenomena, which are one of the most complex cases of material structuring. Among these approaches, phase-field (PF) models (Boettinger et al. 2002; Bragard et al. 2002; Gránácy et al. 2013, 2014) have been successful in describing mesoscale morphologies during polymer crystallization, including dendrites and spherulites. Free-energy functionals in PF modeling can account for crystallization through (i) additional phenomenological order parameters expressing the local fraction of crystalline

phase or (ii) density-dependent terms. It remains to be explored whether some of the available PF formalisms can provide a framework for developing particle-based models describing polymer crystallization.

## 1.3    Defining Operators

Operators transforming a configuration of particle coordinates into an instantaneous field of local order parameters are an integral part of hybrid models. They can be defined through schemes which are either (i) particle-to-mesh (PM) or (ii) based on density distributions.

PM schemes discretize space through a lattice (usually cubic) with $N_{cell}$ nodes. Nevertheless, the coordinates of the CG particles are still distributed in the continuum space underlying the lattice. The operators, used to calculate at each grid point $\mathbf{c}_m$ instantaneous densities or orientation tensors, are defined as:

$$\hat{\rho}_\lambda(\mathbf{c}_m) = C \sum_{i=1}^{n} \sum_{s=1}^{N} \Pi(\mathbf{r}_i(s), \mathbf{c}_m) \gamma_\lambda(i, s) \qquad (9)$$

$$\hat{\mathbf{Q}}_\lambda(\mathbf{c}_m) = \tilde{C} \sum_{i=1}^{n} \sum_{s=1}^{N} \Pi(\mathbf{r}_i(s), \mathbf{c}_m) \mathbf{q}_i(s) \gamma_\lambda(i, s) \qquad (10)$$

Here we assume again for simplicity that the polymer system is comprised of $n$ identical chains which can contain several different monomer species. The chemical type of a CG monomer is defined through $\gamma_\lambda(i, s)$. Namely, $\gamma_\lambda(i, s) = 1$ if the $s$-th segment of the $i$-th chain is of type $\lambda$ and $\gamma_\lambda(i, s) = 0$ otherwise. The function $\Pi(\mathbf{r}_i(s), \mathbf{c}_m)$ defines the rule of how particles with coordinates $\mathbf{r}_i(s)$ are assigned to the grid points. The lattice spacing, $\Delta L$, and the function $\Pi(\mathbf{r}, \mathbf{c})$ are parts of the model definition. As in PM methods in electrostatics (Eastwood et al. 1980), $\Pi(\mathbf{r}_i(s), \mathbf{c}_m)$ is commonly defined (Laradji et al. 1994; Müller and Smith 2005; Daoulas and Müller 2006; Müller 2011; Detcheverry et al. 2008; Qi et al. 2013; Milano and Kawakatsu 2009) through polynomials of $d_\alpha$ ($\alpha = x, y, z$). Here $d_\alpha$ are distances along the three Cartesian directions between the grid point at $\mathbf{c}_m$ and the position of the CG particle. A popular approach is to employ zero-order assignment functions defined as: $\Pi_0 = 1$ if $|d_\alpha| \leq \Delta L/2$ and $\Pi_0 = 0$ otherwise. The normalization prefactors $C$ and $\tilde{C}$ depend on the system. Equation 10 demonstrates that specifying a tensorial operator requires tensors defined on segmental level. In particular, Eq. 10 defines a rule for calculating the local alignment tensor (cf. Eq. 8), where $\mathbf{q}_i(s) = [3\mathbf{u}_i(s) \otimes \mathbf{u}_i(s) - \mathbf{I}]/2$ is a symmetric traceless tensor quantifying the orientation of a characteristic unit vector $\mathbf{u}_i(s)$ assigned to the CG monomer. Instead of $\mathbf{q}_i(s)$ more complex tensors, e.g. biaxial (Gemünden et al. 2013), can be employed to define other tensorial operators. Linking to a coordinate-based representation is not straightforward for all order parameters; PF models describing crystallization are a typical example (Boettinger et al. 2002).

Schemes based on density distributions define operators in continuum. This definition retains translational invariance and isotropicity of space, which are violated in PM schemes because of the lattice. These schemes use the same rules (see Eqs. 9 and 10) as PM methods to express order parameters through particle coordinates, but now the set of node vectors $\mathbf{c}_m$ is replaced by vectors $\mathbf{r}$ defined in continuum space. Instead of $\Pi(\mathbf{r}_i(s), \mathbf{c}_m)$, each particle is assigned a density distribution $W(\mathbf{r}_i(s) - \mathbf{r})$. Frequently, one can give $W(\mathbf{r}_i(s) - \mathbf{r})$ a qualitative physical interpretation, assuming that this mathematical object characterizes the distribution in space of microscopic degrees of freedom that were coarse-grained out. Depending on the problem, the distributions $W(\mathbf{r}_i(s) - \mathbf{r})$ can be isotropic or anisotropic, and their definition is a compromise between physical insights, simplicity of implementation, and computational efficiency.

Two popular families of isotropic distributions are Gaussian (Laradji et al. 1994; Vettorel et al. 2010) and spherical (Müller 2011; Gemünden et al. 2013; Gemünden and Daoulas 2015) functions. The latter are defined as $W(\mathbf{r}) = 3/(4\pi\sigma^3)$ if $|\mathbf{r}| \leq \sigma$ and $W(\mathbf{r}) = 0$ otherwise. The characteristic size of the distribution, $\sigma$, can be chosen (i) empirically (for instance, to achieve a desired amount of interacting neighbors (Pike et al. 2009)) or (ii) to reproduce qualitatively (Vettorel et al. 2010; Gemünden et al. 2013) the dimensions of large groups of microscopic monomers represented by a single CG unit. Ellipsoidal distributions are frequently chosen (Berne and Pechukas 1972; Murat and Kremer 1998; Eurich and Maass 2001) to describe anisotropic CG objects.

## 1.4    Defining Functionals and Parameters: Insights from Mean Field

Field-theoretical approaches based on hybrid models are not the focus of this chapter. Still they deserve some discussion because they often provide the first estimates of how the form and the parameters of the functional affect the properties of the model.

Self-consistent field (SCF) theory is a standard (Edwards 1965; Helfand 1975; Matsen 2006; Müller 2006; Schmid 2011; Fredrickson 2006) method based on field theory. First, the partition function of $n$ interacting chains, Eq. 1, is transformed into a field-theoretical representation, formalized through an equivalent partition function of an ensemble of $n$ independent chains in *fluctuating* fields. This operation is accomplished via the Hubbard-Stratonovich transformation or a method based on insertion of $\delta$-functionals (cf. Eq. 3). Transforming the partition function into a field-theoretical representation shifts the problem of describing the statistical mechanics of mutually interacting molecules to an equally intractable problem of functional integration over fluctuating external fields. SCF theory makes progress by introducing a saddle point approximation for the functional integrals. This operation is equivalent to placing the $n$ independent chains in *average* fields defined through the functional derivative of $H_{nb}$ as:

$$W_\lambda(\mathbf{r}) = \left. \frac{\delta H_{\text{nb}}}{\delta \hat{\rho}_\lambda} \right|_{\hat{\rho}_\lambda(\mathbf{r}) = \bar{\rho}_\lambda(\mathbf{r})} \tag{11}$$

Here $\bar{\rho}_\lambda(\mathbf{r})$ are average densities. In their turn, these average densities are determined by the distribution of the $n$ independent chains in the background fields $W_\lambda(\mathbf{r})$, which closes the self-consistency loop.

Typically the SCF formalism is solved using iterative schemes. When the model of the chain architecture is complex, the average densities at each iteration are calculated over single-chain conformations generated by direct stochastic sampling (Marčelja 1974; Ben-Shaul et al. 1985; Avalos et al. 2004) in the background fields $W_\lambda(\mathbf{r})$. For certain molecular models, the chain statistics in background fields can be obtained more efficiently. For the Gaussian connectivity model, the statistics of chain conformations in $W_\lambda(\mathbf{r})$ is characterized through the probability $q(\mathbf{r}, s)$ of finding a monomer $s$ at point $\mathbf{r}$ of space. This probability is obtained by solving the Edwards diffusion equation (Edwards 1965; Doi and Edwards 1986). The average densities follow from $q(\mathbf{r}, s)$ through a simple integral relationship. For the Gaussian model, several efficient numerical techniques are available (Matsen 2006; Müller 2006; Schmid 2011; Fredrickson 2006) for solving the set of SCF theory equations (including the Edwards equation), such as spectral, pseudo-spectral, and real-space schemes. Over the last years, there has been significant progress (Song et al. 2009; Jiang and Chen 2013) in developing numerical methods for SCF theory based on the wormlike chain model. In this case, the statistics of chain conformations in mean fields follows from the Saito et al differential equation (Saito et al. 1966).

Within SCF theory, free energies can be directly obtained, and one can estimate the thermodynamic properties of a specific model. The accuracy of this estimate depends on the specifics of the system because it is subjected to various approximations inherent to the mean-field SCF theory, such as neglected fluctuations and correlations. Nevertheless, the thermodynamic properties extracted from SCF can assist top-down modeling providing guidelines for choosing parameters and adjusting functionals. SCF results provide the basis for techniques such as the random phase approximation (RPA) which, to some extent, takes into account fluctuations. Structure factors of composition fluctuations in the homogeneous phase derived from RPA are frequently employed in top-down identification of FH parameters (cf. Sect. 1.2).

Frequently, guidelines for top-down modeling are obtained from rudimentary mean-field approximations without engaging the complex machinery of SCF theory. As an illustration, we consider a mixture of homopolymers belonging to several chemical species. There are $n_\lambda$ chains of each homopolymer type comprised of $N_\lambda$ CG monomers. For homogeneous phases, where the composition is described by average densities $\bar{\rho}_\lambda$, a crude mean-field expression for the Helmholtz free energy per volume, $V$, is given by:

$$\frac{\beta F}{V} = \sum_{\lambda} \frac{\bar{\rho}_{\lambda}}{N_{\lambda}} \left[ \ln \left( \frac{\bar{\rho}_{\lambda}}{N_{\lambda} Z_{\lambda}} \right) - 1 \right] + f \left[ \bar{\rho}_{\lambda} \right] \tag{12}$$

The first term of Eq. 12 is the entropy, where $Z_{\lambda}$ stands for the single-molecule partition function of each component. In spirit of the classical FH model, this simple mean-field approximation assumes that $Z_{\lambda}$ do not depend on composition. Hence, only the translational entropy is present. The simple Helmholtz free energy can be employed to estimate observables such as pressure, chemical potential of molecules, and response functions, e.g. isothermal compressibility. Matching chemical potentials in top-down models and simple mean-field expressions obtained from FH-like free energies is often used to fix the degree of monomer incompatibility in the former.

## 1.5    Simulation Techniques for Hybrid Models

Many standard simulation techniques can be straightforwardly applied to hybrid models with PM-based definition of collective variables. In these applications, the grid introduced for the PM definition of the operators also serves to discretize the functional. For example, the partition function in Eq. 1 reads:

$$Z(n, V, T) = \frac{1}{n!} \int d\mathbf{R} \exp \left[ -\sum_{i=1}^{n} \beta H_b(\mathbf{R}_i) \right] \exp \left[ -\underbrace{\sum_{m=1}^{N_{\text{cell}}} f(\{\hat{\rho}_{\lambda}(\mathbf{c}_m)\}) \Delta L^3}_{\beta H_{nb}} \right] \tag{13}$$

where the operators $\hat{\rho}_{\lambda}(\mathbf{c}_m)$ are defined via Eq. 9. For quadratic functionals, the sum over grid cells defining $\beta H_{nb}$ can be rewritten (Daoulas and Müller 2006; Müller 2011) in terms of pairwise potentials between individual particles. These potentials are not invariant translationally. The formalism of pairwise potentials clearly demonstrates (Müller 2011) that the functional in Eq. 13 includes self-interactions.

Partition functions with $\beta H_{nb}$ discretized as in Eq. 13 provide a framework for MC simulations (Laradji et al. 1994; Daoulas and Müller 2006; Müller 2011; Detcheverry et al. 2008). Starting from an initial configuration, standard MC moves, including random segment displacement, slithering snake, chain translation, identity exchange, configuration bias, and chain rebridging algorithms, are used to propose a new configuration. Based on the proposed coordinates of the CG units affected by the MC move, new values of collective coordinates are obtained via the PM operators. The nonbonded energy of the proposed configuration is calculated with the discretized $\beta H_{nb}$. After taking into account the change in the bonded energy (through $\beta H_b$), it is straightforward to formulate a Metropolis criterion for accepting or rejecting the move.

Calculating the changes of collective coordinates during a MC move via a lattice "background" avoids computationally expensive neighbor lists used for nonbonded interactions defined in continuum space. The choice of the assignment function $\Pi(\mathbf{r}_i(s), \mathbf{c}_m)$ affects the computational cost of the MC simulation. Algorithms based on $\Pi_0$ functions are the most efficient and easy to develop, because each particle affects the values of collective coordinates at only one node. It has been demonstrated (Müller 2011) that self-interactions are one of the factors leading to nonuniform distribution of CG particles within one lattice cell. $\Pi_0$ is an exception because self-interactions contribute a constant energy shift which has no effect on the density distribution within a cell (Müller 2011). Formally, the spacing of the grid sets the smallest length scale that can be resolved by a PM model (Daoulas and Müller 2006). Hence, quantifying properties on scales smaller than $\Delta L$ is irrelevant for many applications. One can attempt (Detcheverry et al. 2008; Vogiatzis and Theodorou 2013) to mitigate the breaking of translational invariance by introducing during simulations random lattice displacements. When the displacements are subjected to a Metropolis acceptance criterion (Vogiatzis and Theodorou 2013), the sub-grid structure of the liquid can influence the acceptance rate. With higher-order assignment functions, a single particle contributes to the evaluation of collective coordinates at several grid points. This reduces artifacts related to space discretization, including pinning of interfaces in strongly segregated systems. Nevertheless, the implementation of assignment functions higher than first order is usually undesirable due to substantial increase in computations. One of the exceptions are models of high-molecular-weight polymer melts where molecules are described by chains of linked blobs. Each blob is represented by a Gaussian density cloud distributed over a large number of lattice nodes (Zhang et al. 2013, 2014), i.e., the period of the lattice $\Delta L$ is significantly smaller than the width of the Gaussian. Within this approach, lattice effects become negligible (Zhang et al. 2013) to an extent that one can perform MC simulations in the isobaric ensemble.

MC simulations offer an "exact" (within statistical errors) solution of the statistical mechanics formalized through the partition function in Eq. 13. At the same time, PM techniques provide the framework for approximate but significantly more efficient methods closely linked to field-theoretical approaches. For instance, the "quasi-instantaneous" field approximation presents a powerful concept realized using single-chain-in-mean-field (SCMF) simulations (Müller and Smith 2005; Daoulas and Müller 2006). In SCMF simulations, the instantaneous interactions of a CG monomer with its surroundings are *equivalently* represented through fields defined at each grid node as functional derivatives of $\beta H_{nb}$ with respect to the collective degrees of freedom. Within the "quasi-instantaneous" field approximation, the fields are *frequently* updated substituting the collective coordinates calculated from the coordinates of the CG monomers. Between the updates, the fields remain stationary, and the CG polymer chains evolve in them *independently* typically using local MC moves. The frequent recalculation of the fields recovers partially the correlations between monomers. In the limit where the fields follow instantaneously the configurations of the system, the SCMF simulations become equivalent to the MC simulations. An approach for estimating the accuracy of

the "quasi-instantaneous" field approximation has been elaborated in Daoulas and Müller (2006). PM methods combining particle-based representations with field-based descriptions of interactions have been also employed in context of dynamical algorithms, such as self-consistent Brownian dynamics (SCBD) simulations (Ganesan and Pryamitsyn 2003; Narayanan et al. 2004) and particle-field MD simulations (Milano and Kawakatsu 2009). In dynamical schemes based on hybrid models, forces acting on a particle can be calculated by expressing derivatives of $H_{nb}$ with respect to particle coordinates through the derivatives of the weighting function (Zhang et al. 2017). An alternative approach (Milano and Kawakatsu 2009) is to use stencil operators (Sevink 2015) acting on densities defined on the grid points.

With the help of operators defined through density distributions, nonbonded interactions defined through functionals can be converted into potentials or forces acting in continuum space. This strategy avoids the artifacts of PM schemes but is more demanding computationally, because the interactions must be treated as in standard simulations (including the implementation of cell lists). Quadratic functionals are transformed into pairwise interactions by substituting the definitions of the operators. Namely, scalar and tensorial terms can equivalently be rewritten as:

$$\frac{1}{2} \int d\mathbf{r} \hat{\rho}^2(\mathbf{r}) = \frac{C^2}{2} \sum_{i,s=1}^{n,N} \sum_{j,t=1}^{n,N} \underbrace{\left[ \int d\mathbf{r} W(\mathbf{r}_i(s) - \mathbf{r}) W(\mathbf{r}_j(t) - \mathbf{r}) \right]}_{U(\mathbf{r}_i(s) - \mathbf{r}_j(t))}$$

$$\frac{1}{2} \int d\mathbf{r} \hat{\mathbf{Q}}(\mathbf{r}) : \hat{\mathbf{Q}}(\mathbf{r}) = \frac{\tilde{C}^2}{2} \sum_{i,s=1}^{n,N} \sum_{j,t=1}^{n,N}$$
$$\times \left[ \int d\mathbf{r} W(\mathbf{r}_i(s) - \mathbf{r}) W(\mathbf{r}_j(t) - \mathbf{r}) \right] \mathbf{q}_i(s) : \mathbf{q}_j(t) \qquad (14)$$

For simplicity, we consider here the density of a single component (the subscript $\lambda$ is dropped off) and omit arithmetic as well as material-specific constant prefactors. Only the $1/2$ coefficient is retained to exclude double counting. The integral $U(\mathbf{r}_i(s) - \mathbf{r}_j(t))$ can be calculated analytically for suitable choices of $W(\mathbf{r}_i(s) - \mathbf{r})$ (e.g. the spherical density distribution (Müller 2011) from Sect. 1.3) to obtain explicit pairwise potentials. In simulations of soft matter, it is common to derive CG potentials from overlaps of density distributions, although this approach is not always explicitly linked to density functionals. The anisotropic potential derived analytically from the overlap of two Gaussian ellipsoid density distributions (Berne and Pechukas 1972) is an important example, serving as a basis for developing the standard Gay-Berne potential (Gay and Berne 1981). Models based on overlap of anisotropic objects have been also directly employed (Vink and Schilling 2005; Zong et al. 2012) to study liquid crystalline mesophases. In these models, the LC order is obtained mainly as a consequence of the excluded volume interaction. In contrast, Eq. 14 describes a class of potentials where effects of liquid packing, determined by $U(\mathbf{r}_i(s) - \mathbf{r}_j(t))$, are largely (but not entirely Greco et al. 2016) decoupled

from LC order generated by the tensorial term $\mathbf{q}_i(s){:}\mathbf{q}_j(t)$. This "factorization" offers an easier control over system thermodynamics.

With PM methods, it is straightforward to perform simulations using hybrid models based on functionals with higher order than quadratic (Soga et al. 1995; Daoulas and Müller 2010). However, in this case, it is challenging to use density distributions to derive potentials because high-order polynoms lead to multibody-overlap integrals which cannot be easily calculated analytically. This problem is circumvented in the method of multibody dissipative particle dynamics (DPD) (Pagonabarraga and Frenkel 2001; Warren 2001; Trofimov et al. 2002; Español and Warren 2017). To sketch the approach, we consider a simple system of $n$ identical monomeric particles. The functional of nonbonded interactions is approximately transformed into:

$$\beta H_{nb} = \int d\mathbf{r} f(\hat{\rho}(\mathbf{r})) \simeq \int d\mathbf{r} \sum_{i=1}^{n} \delta(\mathbf{r} - \mathbf{r}_i)\psi(\tilde{\rho}(\mathbf{r})) = \sum_{i=1}^{n} \psi(\tilde{\rho}(\mathbf{r}_i)) \qquad (15)$$

where $\psi(\tilde{\rho}(\mathbf{r}))$ has the meaning of a local free energy per particle (Pagonabarraga and Frenkel 2001). The function $\tilde{\rho}(\mathbf{r})$ is a "smeared-out" local density defined as:

$$\tilde{\rho}(\mathbf{r}) = \int d\mathbf{r}' \sum_{i=1}^{n} w_m(|\mathbf{r} - \mathbf{r}'|)\delta(\mathbf{r}' - \mathbf{r}_i) = \sum_{i=1}^{n} w_m(|\mathbf{r} - \mathbf{r}_i|) \qquad (16)$$

The definition of the "smeared-out" local density is based on the concept of density distributions. However, in multibody DPD, depending on the studied problem, different weighting functions $w_m$ are used to define the smeared-out densities entering the different parts of $\psi(\tilde{\rho}(\mathbf{r}))$. A typical example (Hömberg and Müller 2010) is free energies containing terms which account for short-range repulsions and long-range attractions in the liquid. Differentiating the RHS of Eq. 15 with respect to a particle coordinate, say $\mathbf{r}_i$, generates a force law of *pairwise* forces, which makes simulations straightforward. Obtaining the multibody DPD force law through a density functional presents a "potential-based" derivation. This feature avoids implications related to the "No-go theorem for multibody DPD" (Warren 2013).

## 1.6    Outlook

Modeling polymeric materials by combining particle-based and classical density-functional-based descriptions represents a broad field of research. By defining interaction laws between particles through a free-energy-like functional, these top-down methods enable one to develop systematically particle-based models with desired thermodynamic behavior. Materials described by hybrid models can be studied with standard simulation tools including Monte Carlo sampling and various dynamical methods. In addition to particle-based simulations, hybrid models provide the framework for powerful field-theoretical methods. Although hybrid

modeling has been successful in many fundamental and application-oriented studies of polymeric systems, the field is far from being mature.

One important methodological problem is that the properties of hybrid models cannot be a priori determined in an entirely knowledgeable way even though the interactions are linked to a density functional. The reason is that prior to performing particle-based simulations, the thermodynamic behavior of hybrid models is estimated using approximate mean-field methods. After an initial mean-field-based parameterization, simple hybrid models can be improved using elementary heuristic schemes. However, systematic strategies must be developed in order to refine complex models.

Soft interactions in hybrid models eliminate mechanisms which are crucial for studying polymer dynamics. Currently, there is significant interest in reintroducing the effects of these mechanisms in a top-down fashion. For example, recovering chain non-crossability is necessary to study rheological behavior. Two ideas explored for this purpose are (i) algorithmic detection and prevention of intersection events (Padding and Briels 2011) and (ii) phenomenological interactions (slip-links), temporarily constraining the motion of the chain backbone (Hua and Schieber 1998; Masubuchi et al. 2001; Likhtman 2005; Chappa et al. 2012).

Due to their specific structure, hybrid models have significant potential as platforms for methods where a polymeric system is simultaneously described on particle-based and field-based level. This potential of hybrid models has been already demonstrated through concurrent coupling (Müller 2009) of a continuum Landau-Ginzburg to a hybrid particle-based model. A popular idea is to use hybrid models for semi-discrete representations (Kawakatsu and Kawasaki 1990; Sides et al. 2006; Sevink et al. 2013). These approaches introduce a particle-based description only for some molecular species, while the remaining components are treated on a continuum level. With hybrid models, schemes can be developed (Qi et al. 2013) where the system simultaneously contains regions described on field- and particle-based level. These schemes are field-based analogs of adaptive resolution simulation (AdResS) strategies (Praprotnik et al. 2007; Potestio et al. 2013), which have become very popular in standard particle-based simulations.

# References

Avalos J, Mackie A, Díez-Orrite S (2004) Development of an importance sampling single chain mean field theory for polymer adsorption onto a flat wall. Macromolecules 37:1124–1133

Ben-Shaul A, Szleifer I, Gelbart WM (1985) Chain organization and thermodynamics in micelles and bilayers. I. Theory. J Chem Phys 83:3597–3611

Berne B, Pechukas P (1972) Gaussian model potentials for molecular interactions. J Chem Phys 56:4213–4216

Binder K, Müller M, Virnau P, MacDowell L (2005) Polymer + solvent systems: phase diagrams, interface free energies, and nucleation. Adv Polym Sci 173:1–110

Boettinger W, Warren J, Beckermann C, Karma A (2002) Phase-field simulation of solidification. Annu Rev Mater Res 32:163–194

Bragard J, Karma A, Lee Y, Plapp M (2002) Linking phase-field and atomistic simulations to model dendritic solidification in highly undercooled melts. Interface Sci 10:121–136

Chappa V, Morse D, Zippelius A, Müller M (2012) Translationally invariant slip-spring model for entangled polymer dynamics. Phys Rev Lett 109:148,302

Daoulas KC, Müller M (2006) Single chain in mean field simulations: quasi-instantaneous field approximation and quantitative comparison with Monte Carlo simulations. J Chem Phys 125:184904

Daoulas KC, Müller M (2010) Comparison of simulations of lipid membranes with membranes of block copolymers. Adv Pol Sci 224:197–233

Daoulas KC, Rühle V, Kremer K (2012) Simulations of nematic homopolymer melts using particle-based models with interactions expressed through collective variables. J Phys Condens Matter 24:284121

de Gennes P (1979) Scaling concepts in polymer physics. Cornell University Press, Ithaca/New York

de Gennes P (1980) Dynamics of fluctuations and spinodal decomposition in polymer blends. J Chem Phys 72:4756–4763

de Gennes P, Prost J (1995) The physics of liquid crystals. Oxford University Press, Oxford

Detcheverry F, Kang H, Daoulas KC, Müller M, Nealey P, de Pablo J (2008) Monte Carlo simulations of a coarse grain model for block copolymers and nanocomposites. Macromolecules 41:4989–5001

Doi M, Edwards SF (1986) The theory of polymer dynamics. Oxford University Press, Oxford

Eastwood J, Hockney R, Lowrence D (1980) P3M3DP-the three-dimensional periodic particle-particle/particle-mesh program. Comput Phys Commun 19:215–261

Edwards S (1965) The statistical mechanics of polymers with excluded volume. Proc Phys Soc 85:613–624

Español P, Warren P (2017) Perspective: dissipative particle dynamics. J Chem Phys 146:150901

Eurich F, Maass P (2001) Soft ellipsoid model for Gaussian polymer chains. J Chem Phys 114:7655–7668

Evans R (1979) The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. Adv Phys 28:143–200

Fredrickson GH (2006) The equilibrium theory of inhomogeneous polymers. Clarendon Press, Oxford

Ganesan V, Pryamitsyn V (2003) Dynamical mean-field theory for inhomogeneous polymeric systems. J Chem Phys 118:4345–4348

Gay J, Berne B (1981) Modification of the overlap potential to mimic a linear site-site potential. J Chem Phys 74:3316–3319

Gemünden P, Daoulas KC (2015) Fluctuation spectra in polymer nematics and Frank elastic constants: a coarse-grained modelling study. Soft Matter 11:532–544

Gemünden P, Poelking C, Kremer K, Andrienko D, Daoulas KC (2013) Nematic ordering, conjugation, and density of states of soluble polymeric semiconductors. Macromolecules 46:5762–5774

Gránácy L, Pusztai T, Douglas J (2013) Insights into polymer crystallization from phase-field theory. In: Palsule S (ed) Encyclopedia of polymers and composites. Springer, Berlin/Heidelberg, pp 1–35

Gránácy L, Rátkai L, Szállás A, Korbuly B, Tóth G, Környei L, Pusztai T (2014) Phase-field modeling of polycrystalline solidification: from needle crystals to spherulites – a review. Metall Mater Trans A 45:1694–1719

Greco C, Jiang Y, Chen JZY, Kremer K, Daoulas KC (2016) Maier-Saupe model of polymer nematics: comparing free energies calculated with self consistent field theory and Monte Carlo simulations. J Chem Phys 145:184901

Hansen J, McDonald IR (2006) Theory of simple liquids. Academic Press, Amsterdam

Helfand E (1975) Theory of inhomogeneous polymers: Fundamentals of the Gaussian random-walk model. J Chem Phys 62:999–1005

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871

Hömberg M, Müller M (2010) Main phase transition in lipid bilayers: phase coexistence and line tension in a soft, solvent-free, coarse-grained model. J Chem Phys 132:155104

Hua CC, Schieber JD (1998) Segment connectivity, chain-length breathing, segmental stretch, and constraint release in reptation models. I. Theory and single-step strain predictions. J Chem Phys 109:10018–10027

Jiang Y, Chen J (2013) Self-consistent field theory and numerical scheme for calculating the phase diagram of wormlike diblock copolymers. Phys Rev E 88:042603

Jiang Y, Greco C, Daoulas K, Chen J (2017) Thermodynamics of a compressible Maier-Saupe model based on the self-consistent field theory of wormlike polymer. Polymers 9:48

Kawakatsu T, Kawasaki K (1990) Hybrid models for the dynamics of an immiscible binary mixture with surfactant molecules. Physica A 167:690–735

Kremer K, Müller-Plathe F (2002) Multiscale simulation in polymer science. Mol Simul 28: 729–750

Laradji M, Guo H, Zuckermann MJ (1994) Off-lattice Monte Carlo simulation of polymer brushes in good solvents. Phys Rev E 49:3199–3206

Lifshitz I, Grosberg A, Khokhlov A (1978) Some problems of the statistical physics of polymer chains with volume interaction. Rev Mod Phys 50:683–713

Likhtman A (2005) Single-chain slip-link model of entangled polymers: simultaneous description of neutron spin-echo, rheology, and diffusion. J Chem Phys 38:6128–6139

Lodge T (2003) Block copolymers: past successes and future challenges. Macromol Chem Phys 204:265–273

Löwen H (2002) Density functional theory of inhomogeneous classical fluids: recent developments and new perspectives. J Phys Condens Matter 14:11897–11905

Marčelja S (1974) Chain ordering in liquid crystals II. Structure of bilayer membranes. Biochim Biophys Acta 367:165–176

Masubuchi Y, Takimoto JI, Koyama K, Ianniruberto G, Marrucci G, Greco F (2001) Brownian simulations of a network of reptating primitive chains. J Chem Phys 115:4387–4394

Matsen M (2006) Self-Consistent Field theory and its applications. In: Gompper G, Schick M (eds) Soft Matter, vol 1. Wiley-VCG, Weinheim, pp 87–178

Matteis G, Sonnet A, Virga E (2008) Landau theory for biaxial nematic liquid crystals with two order parameter tensors. Continuum Mech Therm 20:347–374

Mermin ND (1965) Thermal properties of the inhomogeneous electron gas. Phys Rev 137: A1441–A1443

Milano G, Kawakatsu T (2009) Hybrid particle-field molecular dynamics simulations for dense polymer systems. J Chem Phys 130:214106

Morse D, Chung J (2009) On the chain length dependence of local correlations in polymer melts and a perturbation theory of symmetric polymer blends. J Chem Phys 130:224,901

Müller M (2006) Comparison of Self-consistent field theory and Monte Carlo simulations. In: Gompper G, Schick M (eds) Soft matter, vol 1. Wiley-VCG, Weinheim, pp 179–281

Müller M (2009) Concurrent coupling between a particle simulation and a continuum description. Eur Phys J Spec Top 177:149–164

Müller M (2011) Studying amphiphilic self-assembly with soft coarse-grained models. J Stat Phys 145:967–1016

Müller M, Binder K (1995) Computer simulation of asymmetric polymer mixtures. Macro-molecules 28:1825–1834

Müller M, Smith G (2005) Phase separation in binary mixtures containing polymers: a quantitative comparison of single-chain-in-mean-field simulations and computer simulations of the corresponding multichain systems. J Polym Sci, Part B Polym Phys 43:934–958

Müller M, Katsov K, Schick M (2006) Biological and synthetic membranes: what can be learned from a coarse-grained description? Phys Rep 434:113–176

Murat M, Kremer K (1998) From many monomers to many polymers: soft ellipsoid model for polymer melts and mixtures. J Chem Phys 108:4340–4348

Narayanan B, Ganesan V, Pryamitsyn V (2004) Interfacial phenomena in polymer blends: a self-consistent Brownian dynamics study. Macromolecules 37:10180–10194

Noid W, Chu J, Ayton G, Krishna V, Izvekov S, Voth G, Das A, Andersen H (2008) The multiscale coarse graining method. 1. A rigorous bridge between atomistic and coarse-grained models. J Chem Phys 128:244114

Padding J, Briels W (2011) Systematic coarse-graining of the dynamics of entangled polymer melts: the road from chemistry to rheology. J Phys Condens Matter 23:233101

Pagonabarraga I, Frenkel D (2001) Dissipative particle dynamics for interacting systems. J Chem Phys 115:5015–5026

Peter C, Kremer K (2010) Multiscale simulation of soft matter systems. Faraday Discuss 144:9–24

Pike DQ, Detcheverry FA, Müller M, de Pablo JJ (2009) Theoretically informed coarse grain simulations of polymeric systems. J Chem Phys 131:084903

Plischke M, Bergersen B (1994) Equilibrium statistical physics. World Scientific, Singapore

Potestio R, Fritsch S, Español P, Delgado-Buscalioni R, Kremer K, Everaers R, Donadio D (2013) Hamiltonian adaptive resolution simulation for molecular liquids. Phys Rev Lett 110:108301

Praprotnik M, Delle Site L, Kremer K (2007) Multiscale simulation of soft matter: from scale bridging to adaptive resolution. Annu Rev Phys Chem 59:545–571

Prausnitz J, Lichtenthaler R, de Azevedo E (1999) Molecular thermodynamics of fluid-phase equilibria. Prentice Hall PTR, Upper Saddle River

Qi S, Behringer H, Schmid F (2013) Using field theory to construct hybrid particle-continuum simulation schemes with adaptive resolution for soft matter systems. New J Phys 15:125009

Saito N, Takahashi K, Yunoki Y (1966) The statistical mechanical theory of stiff chains. Phys Soc Jpn 22:219–226

Schmid F (2011) Theory and simulation of multiphase polymer systems. In: Boudenne A, Ibos L, Candau Y, Thomas S (eds) Handbook of multiphase polymer systems, vol 1. Wiley-VCG, Weinheim, pp 31–80

Schweizer K, Singh C (1995) Microscopic solubility-parameter theory of polymer blends: general predictions. Macromolecules 28:2063–2080

Sevink G (2015) Rigorous embedding of cell dynamics simulations in the Cahn-Hilliard-Cook framework: Imposing stability and isotropy. Phys Rev E 91:053309

Sevink G, Charlaganov M, Fraaije J (2013) Coarse-grained hybrid simulation of liposomes. Soft Matter 9:2816–2831

Sides S, Kim B, Kramer E, Fredrickson G (2006) Hybrid particle-field simulations of polymer nanocomposites. Phys Rev Lett 96:250601

Soga K, Guo H, Zuckermann M (1995) Polymer brushes in a poor solvent. Europhys Lett 29:531–536

Song W, Tang P, Zhang H, Yang Y, Shi AC (2009) New numerical implementation of self-consistent field theory for semiflexible polymers. Macromolecules 42:6300–6309

Trofimov SY, Nies ELF, Michels MAJ (2002) Thermodynamic consistency in dissipative particle dynamics simulations of strongly nonideal liquids and liquid mixtures. J Chem Phys 117:9383–9394

Vettorel T, Besold G, Kremer K (2010) Fluctuating soft-sphere approach to coarse-graining of polymer models. Soft Matter 6:2282–2292

Vink R, Schilling T (2005) Interfacial tension of the isotropic-nematic interface in suspensions of soft spherocylinders. Phys Rev E 71:051716

Vogiatzis G, Theodorou D (2013) Structure of polymer layers grafted to nanoparticles in silica-polystyrene nanocomposites. Macromolecules 46:4670–4683

Wang ZG (2002) Concentration fluctuation in binary polymer blends: $\chi$ parameter, spinodal and Ginzburg criterion. J Chem Phys 117:481–500

Warren P (2001) Hydrodynamic bubble coarsening in off-critical vapor-liquid phase separation. Phys Rev Lett 87:225702

Warren P (2013) No-go theorem in many-body dissipative particle dynamics. Phys Rev E 87:045303

Wohl K (1946) Thermodynamic evaluation of binary and ternary liquid systems. Trans AIChE 42:215–249

Wu J, Li Z (2007) Density-functional theory for complex fluids. Ann Rev Phys Chem 58:85–112

Zhang G, Daoulas KC, Kremer K (2013) A new coarse grained particle-to-mesh scheme for modeling soft matter. Macromol Chem Phys 214:214–224

Zhang G, Moreira L, Stuehn T, Daoulas KC, Kremer K (2014) Equilibration of high molecular weight polymer melts: a hierarchical strategy. ACS Macro Lett 3:198–203

Zhang S, Qi S, Klushin L, Skvortsov A, Yan D, Schmid F (2017) Anomalous critical slowdown at a first order phase transition in single polymer chains. J Chem Phys 147:064902

Zong J, Zhang X, Wang Q (2012) Fast off-lattice Monte Carlo simulations of soft-core spherocylinders: isotropic-nematic transition and comparisons with virial expansion. J Chem Phys 137:134904

# Multiscale Concepts in Simulations of Organic Semiconductors

# 65

Denis Andrienko

## Contents

**Abstract**

We critically review simulation approaches developed to study organic solar cells, light emitting diodes, and field effect transistors. Special attention is paid to multiscale techniques. In particular, we discuss how to parametrize coarse-grained models for morphology and charge transport simulations, to account for finite-size effects, and to treat long-range interactions in small systems.

## 1  Introduction

Organic semiconductors are conjugated molecular compounds constituted of by carbon and hydrogen atoms and heteroatoms such as nitrogen, sulfur, and oxygen. Examples of organic semiconducting molecules are shown in Fig. 1. Conjugation is the key feature of these materials: Connected $\pi$-orbitals with delocalized electrons can donate or accept electrons easily, thus facilitating charge transfer reactions. Efficient charge transfer enables charge transport on a mesoscale. In their pristine

D. Andrienko (✉)
Max Planck Institute for Polymer Research, Mainz, Germany
e-mail: denis.andrienko@mpip-mainz.mpg.de
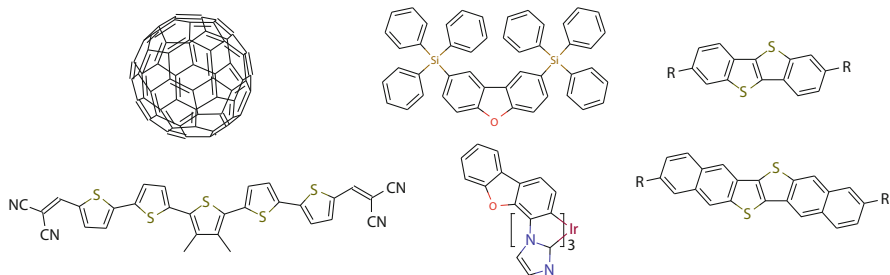
**Fig. 1** Example of a donor-acceptor combination used in an efficient bulk heterojunction solar cell (Fitzner et al. 2011). Host and guest materials forming an active layer of a blue phosphorescent light emitting diode (May et al. 2012a). Self-assembling organic molecules with high charge mobilities measured in field effect transistors (Ebata et al. 2007; Abe et al. 2015)

state, however, organic semiconductors are insulators. They become semiconducting only upon charge injection – from electrodes, by doping, or by photoexcitation.

Besides charge carrier mobility, other relevant physical properties of organic semiconductors are their absorption and emission spectra, ionization potential and electron affinity, solubility, self-assembling abilities, and glass transition temperature. These quantities depend on the molecular arrangement in the solid state, which can vary from crystalline to liquid crystalline to completely amorphous. The prime task of material design is to relate these quantities to the chemical composition and processing conditions, hence helping to prescreen compounds prior to their synthesis. Computer simulations can assist in finding such relationships by providing information which is not accessible to experimental techniques.

In silico prediction of properties of organic semiconducting devices requires multiscale approaches. Indeed, solving the time-dependent Schroedinger or the relativistic Dirac equation is computationally infeasible for an entire device. Likewise, timescales of dynamic processes such as charge or energy transfer span several orders of magnitude. Hence, charge/exciton kinetics cannot be treated by numerical methods with a fixed time step. It is normally coarse-grained into a master equation, as illustrated in Fig. 2, with rates evaluated with the help of the Fermi's golden rule (Athanasopoulos et al. 2007; Kirkpatrick et al. 2007; Nelson et al. 2009; Rühle et al. 2011; Schrader et al. 2012a, b; May et al. 2012a, b; Poelking and Andrienko 2013; Poelking et al. 2013; Kordt et al. 2015a). Multiscaling is essential not only to electronic processes, but also when addressing material processing, since self-assembling properties of conjugated materials, especially polymers can only be studied using coarse-grained models.

In spite of the implicit treatment of the electronic degrees of freedom, the approach based on the master equation is still computationally demanding. Indeed, electronic coupling elements must be evaluated for all pairs of neighboring molecules using first principles calculations. In addition, the evaluation of the self-consistent electrostatic and induction contributions to the free energy of localized charges requires large interaction cutoffs. To reduce the computational overhead,
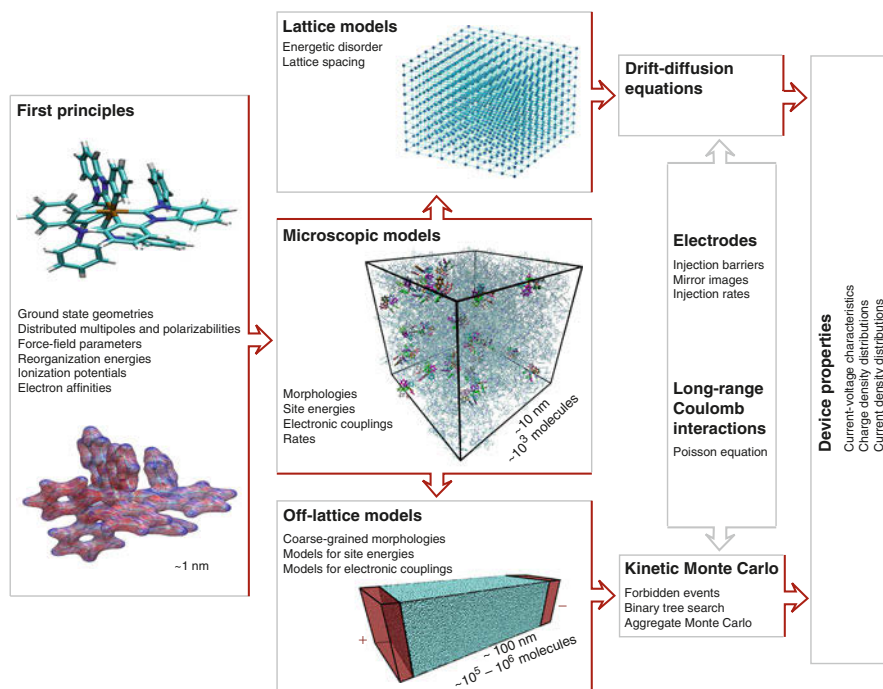
**Fig. 2** Multiscale simulations of organic semiconductors. Polarizable force-fields and electronic properties of isolated molecules obtained from first principles are used to generate morphologies and evaluate charge transfer rates in small systems (microscopic models). Coarse-grained models are parametrized either by matching macroscopic observables, e.g., charge mobility, of the microscopic and coarse-grained (lattice) models. After incorporating long-range electrostatic effects and electrodes, the resulting analytical expressions for mobility are then used to solve drift-diffusion equations for the entire device. Alternatively, off-lattice models can be developed by matching distributions and correlations of site energies, electronic couplings, and positions of molecules. The master equations for this model can be solved using the kinetic Monte Carlo algorithm, yielding macroscopic characteristics of a device. (Reproduced with permission from Adv. Funct. Mater. 25, 1955–1971, 2015)

we can parametrize a lattice model using the solutions to the master equation and then apply this model to study charge/exciton dynamics in larger systems. We will discuss this approach and its pitfalls in Sect. 2.

We can also design a stochastic model, an algorithm which generates the morphologies and the rates of the master equation without computationally demanding evaluations of electronic couplings and site energies. Together with an efficient kinetic Monte Carlo solver, the stochastic model can be used to simulate an entire device. Alternatively, it can be employed as an intermediate model to parameterize the closures of drift-diffusion equations. We will review this approach in Sect. 3.

To further reduce computational costs, we can mimic a macroscopically large system by replicating a relatively small simulation box in two or three spatial

directions. The use of periodic boundary conditions might, however, truncate the tails of the density of states, smoothen the free energy landscape, and artificially increase the charge carrier mobility. These artifacts can be corrected by various extrapolations, which will be discussed in Sect. 4.

In addition, periodic boundary conditions complicate correct treatments of long-range Coulomb forces. To accurately evaluate the density of states, the algorithm should be able to treat charged systems. Unlike the neutral surrounding of the charge, the charge itself should not be periodically repeated. Approaches applicable to partially periodic systems will be discussed in Sect. 5.

## 2    Regular Lattices

From the computational point of view, the most efficient device modeling is based on the drift-diffusion equations (Knapp et al. 2010). Material properties enter these equations through closures, such as dependencies of diffusion constants and mobilities on charge density, temperature, and pressure. Closures provide a link to the chemical composition and material morphology and can be measured experimentally or computed using lower-resolution models. Lattice models are typical minimalist models required to compute such dependencies. In these models, a regular lattice represents material morphology, charge transfer rates decay exponentially with the intermolecular separation, and energy differences obey a Gaussian distribution. This approach was initiated by Bässler and has been used by several groups to understand the role of traps, finite charge carrier density, energetic disorder, and other mesoscopic parameters on charge mobility (Pasveer et al. 2005; Cottaar et al. 2011; van der Holst et al. 2011; Bässler 1993; Yimer et al. 2009; Novikov et al. 1998; Freire and Tonezer 2009).

Fits of phenomenological expressions for charge carrier mobility to the kinetic Monte Carlo simulations of lattice models are known as extended (correlated) Gaussian disorder models, EGDM, and ECDM. These expressions depend parametrically on the lattice constant $a$, the energetic disorder $\sigma$, and a prefactor $\mu_0$. In principle, $a$ and $\sigma$ can be evaluated in a relatively small system: $a$ as the mean distance between neighboring molecules and $\sigma$ as the width of the DOS that results from perturbative energy calculations. $\mu_0$ can be extracted from charge transport simulations performed at different temperatures. This approach, however, does not lead to reliable parameterizations (Kordt et al. 2014, 2015a). Indeed, a multidimensional fit of simulated mobilities to the EGDM or ECDM expressions yields a very different set of parameters: the EGDM underestimates the energetic disorder, while the ECDM overestimates it. In both cases spatial site energy correlations are responsible for this discrepancy. EGDM does not include correlations and compensates for higher mobility values by reducing the energetic disorder $\sigma$. ECDM overestimates spatial correlations and compensates this by reducing the lattice constant (Kordt and Andrienko 2016). The discrepancy between microscopic values and fits to EGDM and ECDM tells us that parameters of these models do not have a clear physical interpretation. Nevertheless, they still provide

reasonable parameterizations and can eventually be used in conjunction with drift-diffusion equations.

A more conceptual drawback of EGDM and ECDM resides in the approximations incorporated in the rates of the master equation of the underlying lattice model, which constrain the functional form of $\mu(\rho, \mathbf{F}, T)$. To avoid such constraints, we can simply tabulate the mobility in a wide range of charge densities, temperatures, and electric fields with the help of the microscopic model. The tabulated function can then be used in the drift-diffusion equations solver (Kordt et al. 2015b). Computationally, the tabulation is only feasible with the help of a stochastic model, which we will discuss in the next section.

## 3    Stochastic Models

In heterogeneous organic semiconducting devices, charges are nonuniformly distributed and charge density variations span several orders of magnitude. To cover the required density range in simulations, we need to simulate relatively large systems. This quickly becomes computationally demanding if all rates are evaluated from first principles (Kordt et al. 2015b). To remedy the situation, we can devise a phenomenological algorithm to parametrize the master equation (Baumeier et al. 2012; Stenzel et al. 2014; Brereton et al. 2014; Stenzel et al. 2014).

In case of morphologies, this algorithm reproduces given correlation functions, densities, and coordination numbers. For (approximately) spherically shaped molecules, the pair correlation function, or radial distribution function, $g(r)$, contains the most relevant structural information. To reproduce this function approximately, we can use a Poisson process (Baumeier et al. 2012; Kordt et al. 2014). More accurate coarse-graining techniques, such as iterative Boltzmann inversion (Tschoep et al. 1998; Reith et al. 2003) or inverse Monte Carlo (Lyubartsev and Laaksonen 1995; Soper 1996; Murtola et al. 2009), allow an exact reproduction of the radial distribution function (Kordt et al. 2015a). These methods optimize a pair interaction potential, $U(r)$, in a way that the corresponding $g(r)$ is reproduced. The approach can also be applied to non-spherical molecules, by using several interacting sites per molecule (Rühle et al. 2009).

The second ingredient of the stochastic model is the connectivity. In the atomistic model only molecules within a certain cutoff distance are used for calculating charge transfer rates, while the rest of the rates is set to zero. This is justified by the fact that electronic coupling elements decrease roughly exponentially with molecular separation (Kordt et al. 2015a). The distance that determines whether or not two molecules are connected is given by their two closest atoms. Since this information is not present in the coarse-grained model, the resulting probability of two sites being connected is given by the corresponding probability extracted from the atomistic system.

For charge transfer rates, a simple procedure can be developed by analyzing the distributions of parameters entering the charge transfer rate. For example, in an amorphous mesophase of the organic semiconductor $Alq_3$, the distribution

of electronic couplings at every particular intermolecular separation is Gaussian. The dependence of the width and the mean of this Gaussian on intermolecular separation can be determined from microscopic simulations. The spatial correlations of site energies can be introduced by using a moving-average procedure. In this approach, the site energies of the neighbors within a certain cutoff are mixed into initially independent Gaussian site-energy distribution (Kordt and Andrienko 2016). Electronic coupling elements can also be generated using appropriate distributions. These distributions are, however, separation-dependent: the logarithm of squared transfer integrals, $\log J^2$ (which is often Gaussian distributed) depends on molecular separation. In the stochastic model, transfer integrals are then drawn from the distant-dependent distributions.

Stochastic models developed for amorphous mesophases of Alq$_3$ and DCV4T could reproduce the mobility-field and mobility-density dependencies (Baumeier et al. 2012; Stenzel et al. 2014; Kordt et al. 2014). This indicates that they indeed can serve as an intermediate step between the completely microscopic descriptions and macroscopic, drift-diffusion-equation-based models. The current challenge is to extend such models to anisotropic and heterogeneous systems.

## 4 Finite-Size Effects

The dependence of macroscopic observables on the system size has been experimentally observed in organic semiconductors in systems with large energetic disorder or broad density of states. For example, time-of-flight measurements of mobility are impossible in thin organic films, where charge transport is dispersive and transients do not have a characteristic plateau used to determine the transient time (Scher and Montroll 1975; Borsenberger et al. 1991, 1992, 1993). Using thicker samples normally remedies the situation.

A similar effect is also observed in simulations of small systems. It turns out that replicating the simulation box still gives incorrect (higher) values of charge carrier mobility (Lukyanov and Andrienko 2010). Indeed, all duplicated boxes have exactly the same (and small) number of independent site energies. Statistical averages are performed over this small set of site energies, and charge carriers traverse the sample at a different (higher) temperature than in an infinitely large system. Surprisingly, an additional averaging over different, statistically uncorrelated, but still repeated snapshots does not help, since the origin of the problem lies in the limited number of available for every distinct Monte Carlo simulation sites.

This type of finite-size effects becomes much smaller at higher temperatures, since the relevant dimensionless parameter (at least in the Gaussian disorder model) is the width of the site energy distribution $\sigma$ divided by $k_B T$. In fact, an empirical expression, $(\sigma/k_B T_{ND})^2 = -5.7 + 1.05 \ln N$, can be used to estimate at what temperature the "transition" between the dispersive and non-dispersive transport occurs, where $N$ is the number of the hopping sites (molecules) in the system. This observation has been used to perform an empirical correction of finite-size effects (Lukyanov and Andrienko 2010): Nondispersive mobilities were calculated

for a set of temperatures above $T_{\mathrm{ND}}$. Then, an explicit temperature dependence was used to extrapolate the value of mobility to room temperature.

This method relies on an explicit knowledge of the temperature dependence of mobility. While the exact analytical expression of this dependence is available only for one-dimensional systems (Derrida 1983; Seki and Tachiya 2001),

$$\mu(T) = \frac{\mu_0}{T^{3/2}} \exp\left[ -\left(\frac{a}{T}\right)^2 - \left(\frac{b}{T}\right) \right],$$

it can still be used in a three-dimensional case in a rather broad temperature range, as was verified by performing simulations for systems of different sizes and at different temperatures (Lukyanov and Andrienko 2010).

An alternative method is to determine the mobility as a function of the box size. For one-dimensional systems and Gaussian DOS, the mobility is inversely proportional to the number of visited sites $N$ (Kordt et al. 2016),

$$\mu(N) = \mu_\infty \left(1 + \frac{c}{N}\right).$$

Using this relation, it is possible to extrapolate the finite-size mobilities to its macroscopic value $\mu_\infty$.

To illustrate the relevance of such extrapolation, charge transport in the amorphous mesophase of $Alq_3$ was simulated in systems of different sizes (Lukyanov and Andrienko 2010). In a system of 512 molecules, the simulated mobility was of the order of $10^{-6}\,\mathrm{cm^2/Vs}$, while in a box of 4096 molecules, an order of magnitude lower value was measured. Extrapolation procedure resulted in a mobility of $10^{-9}\,\mathrm{cm^2/Vs}$, which is three orders of magnitude lower than the one simulated in a small system. Note that the magnitude of the correction is very sensitive to the value of the energetic disorder $\sigma$. While in $Alq_3$ $\sigma = 0.14\,\mathrm{eV}$, in an amorphous DCV4T $\sigma = 0.25\,\mathrm{eV}$, the value of mobility can be overestimated by seven(!) orders of magnitude (Kordt et al. 2014).

## 5   Long-Range Interactions

An important ingredient of charge and energy transfer reactions is the free energy difference between the final and initial states or the driving force. In organic semiconductors this energy difference is often evaluated in a perturbative way, starting from an isolated molecule and then calculating the electrostatic and induction contributions of the environment. Coulomb interactions of partial charges or higher distributed multipoles are inherently long-range and require special summation techniques, e.g., Ewald summation, which is widely used in atomistic molecular dynamics simulations (Ewald 1921). In this approach, the Coulomb potential is split into two terms, one of which is converging fast in real and the other in reciprocal space. Induction contribution can also be incorporated in such a

scheme (Ren and Ponder 2003). While this method is well developed, it is designed for neutral systems, whereas for charge transport, our interest lies in energies of a localized charge interacting with the neutral environment. The presence of a charged excitation and its polarization cloud violates the periodicity of the system, which is essential for calculations in the reciprocal space.

A computationally more demanding solution is to evaluate electrostatic sums for spheres centered around the charged molecule (bulk) or cylinders (slab) of radius $R$. In both cases charge-quadrupole interactions scale as $1/R$, making it possible to extrapolate to $R \rightarrow \infty$ (D'Avino et al. 2016). A more elegant solution to this problem has been proposed by Poelking and Andrienko (2016). The nonperiodic (foreground) part of the system which incorporates the charged excitation and its induction cloud is superimposed onto a periodic, neutral background, which is computed using the Ewald summation method. The real-space interaction between these two regions is mediated by fields created by the background charge distribution, including induced moments. A modified shape term (Smith 1981) is added to account for the net charge and quadrupole of the simulation box. This term takes into account surface effects and depends on the summation geometry.

By applying this technique to organic/organic interfaces, it has been shown that a cutoff of $4-8$ nm is sufficient to converge the energy of a periodic three-dimensional system (Schwarze et al. 2016). For ordered interfaces, the convergence turns out to be significantly slower: for a 12 nm thick slab, the energy is far from converged even for the cutoff of 22 nm (Poelking and Andrienko 2016, 2015; Poelking et al. 2014). Hence, this method should be used for two-dimensional periodic systems or, in general, heterogeneous systems with a long-range molecular ordering. The differences between using simulations with a cutoff and without it are remarkable: the additional electrostatic/induction contribution can change from 0.5 eV (10 nm cutoff) to $-0.7$ eV (infinite system). The correct treatment of electrostatic can therefore *reverse* the role of donor and acceptor as conditioned by gas-phase energy levels and energy levels calculated with a seemingly ample cutoff. It also predicts that the energy profiles for electrons and holes are flat at the organic-organic interfaces, while cutoff-based calculations lead to a significant level bending. Finally, it allows to establish a relationship between structural coherence and state energetics: the structural coherence is probed up to a $\mu m$ scale, hence emphasizing the role of extended crystallites (and their alignment) at interfaces.

## 6    Outlook

There are still a few directions where substantial method development is required to achieve a parameter-free modeling of realistic devices. An important step is to explicitly include the many-body induction interactions to the solution of the master equation. This, however, requires the reevaluation of rates at every Monte Carlo

step and is computationally demanding. Another issue is the quantitative treatment of excited states embedded in a heterogeneous polarizable molecular environment, including descriptions of charge-exciton and exciton-exciton interactions. These advancements are absolutely vital for devising accurate structure-property relationships for organic semiconductors.

# References

Abe M, Mori T, Osaka I, Sugimoto K, Takimiya K (2015) Thermally, operationally, and environmentally stable organic thin-film transistors based on Bis[1]benzothieno[2,3-d:2',3'-d']naphtho[2,3-b:6,7-b']dithiophene derivatives: effective synthesis, electronic structures, and structure–property relationship. Chem Mater 27(14):5049–5057. https://doi.org/10.1021/acs.chemmater.5b01608

Athanasopoulos S, Kirkpatrick J, Martínez D, Frost JM, Foden CM, Walker AB, Nelson J (2007) Predictive study of charge transport in disordered semiconducting polymers. Nano Lett 7(6):1785–1788. https://doi.org/10.1021/nl0708718

Bässler H (1993) Charge transport in disordered organic photoconductors: a Monte Carlo simulation study. Physica Status Solidi (b) 175(1):15–56. https://doi.org/10.1002/pssb.2221750102

Baumeier B, Stenzel O, Poelking C, Andrienko D, Schmidt V (2012) Stochastic modeling of molecular charge transport networks. Phys Rev B 86(18):184202. https://doi.org/10.1103/PhysRevB.86.184202

Borsenberger P, Pautmeier L, Bässler H (1991) Charge transport in disordered molecular solids. J Chem Phys 94(8):5447. https://doi.org/10.1063/1.460506, http://link.aip.org/link/JCPSA6/v94/i8/p5447/s1&Agg=doi

Borsenberger P, Pautmeier L, Bässler H (1992) Nondispersive-to-dispersive charge-transport transition in disordered molecular solids. Phys Rev B 46(19):12145–12153. https://doi.org/10.1103/PhysRevB.46.12145, http://prb.aps.org/abstract/PRB/v46/i19/p12145_1

Borsenberger PM, Magin EH, Van Auweraer MD, De Schryver FC (1993) The role of disorder on charge transport in molecularly doped polymers and related materials. Physica Status Solidi (a) 140(1):9–47. https://doi.org/10.1002/pssa.2211400102, http://onlinelibrary.wiley.com/doi/10.1002/pssa.2211400102/abstract

Brereton T, Stenzel O, Baumeier B, Andrienko D, Schmidt V, Kroese D (2014) Efficient simulation of Markov Chains using segmentation. Methodol Comput Appl Probab 16(2):465–484. https://doi.org/10.1007/s11009-013-9327-x

Cottaar J, Koster LJA, Coehoorn R, Bobbert PA (2011) Scaling theory for percolative charge transport in disordered molecular semiconductors. Phys Rev Lett 107(13):136601. https://doi.org/10.1103/PhysRevLett.107.136601

D'Avino G, Muccioli L, Castet F, Poelking C, Andrienko D, Soos ZG, Jérôme Cornil, Beljonne D (2016) Electrostatic phenomena in organic semiconductors: fundamentals and implications for photovoltaics. J Phys Condens Matter 28(43):433002. https://doi.org/10.1088/0953-8984/28/43/433002, http://stacks.iop.org/0953-8984/28/i=43/a=433002

Derrida B (1983) Velocity and diffusion constant of a periodic one-dimensional hopping model. J Stat Phys 31(3):433–450. https://doi.org/10.1007/BF01019492

Ebata H, Izawa T, Miyazaki E, Takimiya K, Ikeda M, Kuwabara H, Yui T (2007) Highly soluble [1]Benzothieno[3,2-b]benzothiophene (BTBT) derivatives for high-performance, solution-processed organic field-effect transistors. J Am Chem Soc 129(51):15732–15733. https://doi.org/10.1021/ja074841i

Ewald PP (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale. Annalen der Physik 369(3):253–287. https://doi.org/10.1002/andp.19213690304

Fitzner R, Reinold E, Mishra A, Mena-Osteritz E, Ziehlke H, Körner C, Leo K, Riede M, Weil M, Tsaryova O, Weiß A, Uhrich C, Pfeiffer M, Bäuerle P (2011) Dicyanovinyl–Substituted Oligothiophenes: structure-property relationships and application in vacuum-processed small molecule organic solar cells. Adv Funct Mater 21(5):897–910. https://doi.org/10.1002/adfm.201001639

Freire JA, Tonezer C (2009) Density of states and energetic correlation in disordered molecular systems due to induced dipoles. J Chem Phys 130(13):134901. https://doi.org/10.1063/1.3090482

Kirkpatrick J, Marcon V, Nelson J, Kremer K, Andrienko D (2007) Charge mobility of discotic mesophases: a multiscale quantum and classical study. Phys Rev Lett 98(22):227402. https://doi.org/10.1103/PhysRevLett.98.227402

Knapp E, Häusermann R, Schwarzenbach HU, Ruhstaller B (2010) Numerical simulation of charge transport in disordered organic semiconductor devices. J Appl Phys 108(5):054504. https://doi.org/10.1063/1.3475505

Kordt P, Andrienko D (2016) Modeling of spatially correlated energetic disorder in organic semiconductors. J Chem Theory Comput 12(1):36–40. https://doi.org/10.1021/acs.jctc.5b00764

Kordt P, Stenzel O, Baumeier B, Schmidt V, Andrienko D (2014) Parametrization of extended Gaussian disorder models from microscopic charge transport simulations. J Chem Theory Comput 10(6):2508–2513. https://doi.org/10.1021/ct500269r

Kordt P, van der Holst JJM, Al Helwi M, Kowalsky W, May F, Badinski A, Lennartz C, Andrienko D (2015a) Modeling of organic light emitting diodes: from molecular to device properties. Adv Funct Mater 25(13):1955–1971. https://doi.org/10.1002/adfm.201403004

Kordt P, Stodtmann S, Badinski A, Helwi MA, Lennartz C, Andrienko D (2015b) Parameter-free continuous drift–diffusion models of amorphous organic semiconductors. Phys Chem Chem Phys 17(35):22778–22783. https://doi.org/10.1039/C5CP03605D, http://pubs.rsc.org/en/content/articlelanding/2015/cp/c5cp03605d

Kordt P, Speck T, Andrienko D (2016) Finite-size scaling of charge carrier mobility in disordered organic semiconductors. Phys Rev B 94:014208. https://doi.org/10.1103/PhysRevB.94.014208

Lukyanov A, Andrienko D (2010) Extracting nondispersive charge carrier mobilities of organic semiconductors from simulations of small systems. Phys Rev B 82(19):193202. https://doi.org/10.1103/PhysRevB.82.193202

Lyubartsev AP, Laaksonen A (1995) Calculation of effective interaction potentials from radial distribution functions: a reverse Monte Carlo approach. Phys Rev E 52(4):3730–3737. https://doi.org/10.1103/PhysRevE.52.3730

May F, Al-Helwi M, Baumeier B, Kowalsky W, Fuchs E, Lennartz C, Andrienko D (2012a) Design rules for charge-transport efficient host materials for phosphorescent organic light-emitting diodes. J Am Chem Soc 134(33):13818–13822. https://doi.org/10.1021/ja305310r

May F, Baumeier B, Lennartz C, Andrienko D (2012b) Can lattice models predict the density of states of amorphous organic semiconductors? Phys Rev Lett 109(13):136401. https://doi.org/10.1103/PhysRevLett.109.136401

Murtola T, Bunker A, Vattulainen I, Deserno M, Karttunen M (2009) Multiscale modeling of emergent materials: biological and soft matter. Phys Chem Chem Phys 11(12):1869–1892. https://doi.org/10.1039/B818051B, http://pubs.rsc.org/en/content/articlelanding/2009/cp/b818051b

Nelson J, Kwiatkowski JJ, Kirkpatrick J, Frost JM (2009) Modeling charge transport in organic photovoltaic materials. Acc Chem Res 42(11):1768–1778. https://doi.org/10.1021/ar900119f

Novikov SV, Dunlap DH, Kenkre VM, Parris PE, Vannikov AV (1998) Essential role of correlations in governing charge transport in disordered organic materials. Phys Rev Lett 81(20):4472–4475. https://doi.org/10.1103/PhysRevLett.81.4472

Pasveer WF, Cottaar J, Tanase C, Coehoorn R, Bobbert PA, Blom PWM, de Leeuw DM, Michels MAJ (2005) Unified description of charge-carrier mobilities in disordered semiconducting polymers. Phys Rev Lett 94(20):206601. https://doi.org/10.1103/PhysRevLett.94.206601

Poelking C, Andrienko D (2013) Effect of polymorphism, regioregularity and paracrystallinity on charge transport in poly(3-hexylthiophene) [P3ht] nanofibers. Macromolecules 46(22): 8941–8956. https://doi.org/10.1021/ma4015966

Poelking C, Andrienko D (2015) Design rules for organic donor-acceptor heterojunctions: pathway for charge splitting and detrapping. J Am Chem Soc 6320–6326. https://doi.org/10.1021/jacs.5b02130

Poelking C, Andrienko D (2016) Long-range embedding of molecular ions and excitations in a polarizable molecular environment. J Chem Theory Comput 12(9):4516–4523. https://doi.org/10.1021/acs.jctc.6b00599

Poelking C, Cho E, Malafeev A, Ivanov V, Kremer K, Risko C, Brédas JL, Andrienko D (2013) Characterization of charge-carrier transport in semicrystalline polymers: electronic couplings, site energies, and charge-carrier dynamics in poly(bithiophene-*alt*-thienothiophene) [PBTTT]. J Phys Chem C 117(4):1633–1640. https://doi.org/10.1021/jp311160y

Poelking C, Tietze M, Elschner C, Olthof S, Hertel D, Baumeier B, Würthner F, Meerholz K, Leo K, Andrienko D (2014) Impact of mesoscale order on open-circuit voltage in organic solar cells. Nat Mater 14(4):434–439. http://www.nature.com/doifinder/10.1038/nmat4167

Reith D, Pütz M, Müller-Plathe F (2003) Deriving effective mesoscale potentials from atomistic simulations. J Comput Chem 24(13):1624–1636. https://doi.org/10.1002/jcc.10307

Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107(24):5933–5947. https://doi.org/10.1021/jp027815+

Rühle V, Junghans C, Lukyanov A, Kremer K, Andrienko D (2009) Versatile object-oriented toolkit for coarse-graining applications. J Chem Theory Comput 5(12):3211–3223. https://doi.org/10.1021/ct900369w

Rühle V, Lukyanov A, May F, Schrader M, Vehoff T, Kirkpatrick J, Baumeier B, Andrienko D (2011) Microscopic simulations of charge transport in disordered organic semiconductors. J Chem Theory Comput 7(10):3335–3345. https://doi.org/10.1021/ct200388s

Scher H, Montroll EW (1975) Anomalous transit-time dispersion in amorphous solids. Phys Rev B 12(6):2455–2477. https://doi.org/10.1103/PhysRevB.12.2455

Schrader M, Fitzner R, Hein M, Elschner C, Baumeier B, Leo K, Riede M, Bäuerle P, Andrienko D (2012a) Comparative study of microscopic charge dynamics in crystalline acceptor-substituted oligothiophenes. J Am Chem Soc 134(13):6052–6056. https://doi.org/10.1021/ja300851q

Schrader M, Körner C, Elschner C, Andrienko D (2012b) Charge transport in amorphous and smectic mesophases of dicyanovinyl-substituted oligothiophenes. J Mater Chem 22(41): 22258–22264. https://doi.org/10.1039/C2JM34837C, http://pubs.rsc.org/en/content/articleland ing/2012/jm/c2jm34837c

Schwarze M, Tress W, Beyer B, Gao F, Scholz R, Poelking C, Ortstein K, Günther AA, Kasemann D, Andrienko D, Leo K (2016) Band structure engineering in organic semiconductors. Science 352(6292):1446–1449. https://doi.org/10.1126/science.aaf0590, http://science.sciencemag.org/content/352/6292/1446

Seki K, Tachiya M (2001) Electric field dependence of charge mobility in energetically disordered materials: polaron aspects. Phys Rev B 65:014305. https://doi.org/10.1103/PhysRevB.65.014305

Smith ER (1981) Electrostatic energy in ionic crystals. Proc R Soc London A Math Phys Sci 375(1763):475–505

Soper AK (1996) Empirical potential Monte Carlo simulation of fluid structure. Chem Phys 202(2–3):295–306. https://doi.org/10.1016/0301-0104(95)00357-6, http://www.sciencedirect.com/science/article/pii/0301010495003576

Stenzel O, Hirsch C, Brereton T, Baumeier B, Andrienko D, Kroese D, Schmidt V (2014) A general framework for consistent estimation of charge transport properties via random walks in random environments. Multiscale Model Simul 12(3):1108–1134. https://doi.org/10.1137/130942504

Tschoep W, Kremer K, Batoulis J, Buerger T, Hahn O (1998) Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. Acta Polym 49(2–3):61–74. https://doi.org/10.1002/(SICI)1521-4044(199802)49:2/3<61::AID-APOL61>3.0.CO;2-V

van der Holst JJM, van Oost FWA, Coehoorn R, Bobbert PA (2011) Monte Carlo study of charge transport in organic sandwich-type single-carrier devices: effects of Coulomb interactions. Phys Rev B 83(8):085206. https://doi.org/10.1103/PhysRevB.83.085206

Yimer Y, Bobbert P, Coehoorn R (2009) Charge transport in disordered organic host–guest systems: effects of carrier density and electric field. Synth Met 159(21–22):2399–2401. https://doi.org/10.1016/j.synthmet.2009.10.025, http://www.sciencedirect.com/science/article/pii/S037967790900544X

# Adaptive Resolution Molecular Dynamics Technique

# 66

## M. Praprotnik, R. Cortes-Huerto, R. Potestio, and L. Delle Site

## Contents

### Abstract

Soft matter systems display properties that span different time and length scales. In addition, scales' interplay is often the key to understand fundamental mechanisms to the aim of controlling and/or designing materials with properties on demand. On the other hand, computational soft matter is limited by computational power for both, size and time of simulation and analysis of large sets of

M. Praprotnik (✉)
Laboratory for Molecular Modeling, National Institute of Chemistry,
Ljubljana, Slovenia
Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana,
Slovenia
e-mail: praprot@cmm.ki.si

R. Cortes-Huerto · R. Potestio
Max Planck Institute for Polymer Research, Mainz, Germany
e-mail: corteshu@mpip-mainz.mpg.de; raffaello.potestio@unitn.it

L. Delle Site
Institute for Mathematics, Freie Universitat Berlin, Berlin, Germany
e-mail: luigi.dellesite@fu-berlin.de

data. In this perspective, computational efficiency to treat large systems on long time scales becomes one of the main goals in constructing modern algorithms, together with the capability of designing theoretical schemes for data analysis capable of extracting the relevant information of interest above all the effects of scales' interplay. One common and recurrent feature, in such studies, is the need to include relevant chemical details in a specific region where an event of interest is taking place, while the environment plays simply the role of a macroscopic thermodynamic bath that can be treatable at a coarse-grained level. Thus, an efficient computational strategy consists in employing multiple resolution methods, which simultaneously consider models with different resolution in different regions. This chapter provides a basic introduction to the adaptive resolution simulation (AdResS) method and its recent extensions. This methodology is designed with the idea of efficient computation and analysis of multiple scales as envisaged above. We will report its basic principles and technical aspects for the various directions along which the original idea was developed. As it will emerge in the next sections, the basic idea of adaptive resolution, already highly efficient in its first implementation, has now reached a high level of theoretical solidity, being framed in different but complementary ways in physically rigorous principles. Finally, selected applications, relevant in the field of materials science, chemical physics, and biochemistry, are illustrated in order to show the advanced possibilities of application of the method.

# 1    Adaptive Resolution Simulations

AdResS (Praprotnik et al. 2005, 2008) is a multi-resolution simulation method that links two regions of a simulation box having different resolutions: region 1, atomistic (AT), and region 2, coarse-grained (CG). The innovative aspect of such a setup is that the boundary between such regions is open; thus, it allows the free exchange of particles/molecules between the two regions. This free exchange occurs through a coupling transition region (HY) where molecules have space-dependent hybrid atomistic/coarse-grained resolution or intermolecular interactions. In the original version (Praprotnik et al. 2005, 2008), the guiding principle employed was that the coupling between two regions must be smooth enough so that the dynamics of the atomistic and the dynamics of the coarse-grained region, through a hybrid resolution region, would not be perturbed in a sizable way, i.e., Newton's 3rd law should be fulfilled. This empirical principle corresponds to a straightforward molecular dynamics (MD) algorithm based on the interpolation of the atomistic and coarse-grained forces. The above requirement prevents the use of an interpolation based on a Hamiltonian (Delle Site 2007). While for most typical applications not of a serious concern and for some even advantageous, e.g., hydrodynamics, the desire to base such an interpolation scheme on a Hamiltonian description led to the development of H-AdResS (Potestio et al. 2013a, b). At the cost of a weak violation of Newton's 3rd law, the Hamiltonian description offers some advantages, including the option to perform adaptive Monte Carlo (MC) simulations. Recently,

AdResS has been also extended to simulate open molecular systems within the grand canonical ensemble by grand canonical AdResS (GC-AdResS) (Wang et al. 2012, 2013) and open boundary molecular dynamics (OBMD) (Delgado-Buscalioni et al. 2015; Sablić et al. 2016, 2017a, b). In the following, the different adaptive resolution approaches will be introduced, and some typical applications will be presented.

## 2 Adaptive Resolution Simulations: AdResS

The original concept is based on the idea that forces between particles/molecules should change upon variation of the molecular resolution and that this should, in principle, not affect the intermolecular dynamics. Thus, the explicit coupling is such that the force between two molecules $\alpha$ and $\beta$ is computed via a space-dependent interpolation formula and is written as

$$F_{\alpha\beta} = w(X_\alpha)w(X_\alpha)F_{\alpha\beta}^{AT} + [1 - w(X_\alpha)w(X_\alpha)]F_{\alpha\beta}^{CG} \tag{1}$$

here, $F_{\alpha\beta}^{AT}$ is the atomistic force between the particles and $F_{\alpha\beta}^{CG}$ is the coarse-grained force; $w(x)$ is the interpolating function and is defined as

$$\begin{Bmatrix} 1 & x < d_{AT} \\ cos^2[\pi/(2d_\Delta)(x - d_{AT})] & d_{AT} < x < d_{AT} + d_\Delta \\ 0 & d_{AT} + d_\Delta < x \end{Bmatrix}$$

where $d_{AT}$ and $d_\Delta$ are the linear dimension of the atomistic region and hybrid regions, respectively (see Fig. 1). The weighting function smoothly goes from 0 to 1 in the hybrid region and allows the coarse-grained particles to change their resolution into an atomistic molecule and vice versa. A local thermostat takes care of adsorbing and releasing the excess of heat produced by the change of resolution (see e.g., Poblete et al. 2010). This minimal ansatz has been shown to be already sufficient for performing accurate simulations in both the atomistic and coarse-grained regions (see e.g., Praprotnik et al. 2007b). Other more natural choices of coupling, e.g., through an energy, were not considered at this point. Provided that the dynamics should be governed by intermolecular forces only, it was shown that a spatial interpolation of any energy functional, instead of forces, leads to the mathematical nonexistence of a global physically consistent energy in the adaptive sense. As a consequence, any energy functional constructed on the basis of spatial interpolation would be artificial from the physical point of view (see Eqs. 14 and 15 in Delle Site 2007). Nevertheless, relaxing the constraint on the exact conservation of Newton's 3rd law, energy-based algorithms, which have the technical capability of conserving energy (a familiar concept for practitioners of MD and needed for MC) and enable more flexibility in the choice of the standard technical setups, can be devised. In both types of approach one needs to carefully consider the
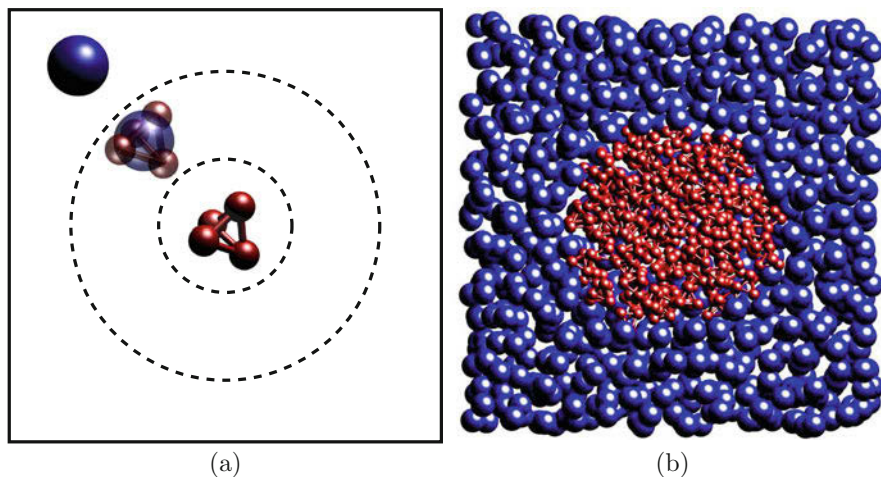
**Fig. 1** (**a**) Cartoon of the AdResS setup with spherical regions modeled with different levels of resolution. In the center, there is the AT domain with the molecules surrounded by the HY and CG domains, respectively. (**b**) Multiscale system of a liquid composed of tetrahedral molecules. (Reprinted from Praprotnik et al. 2006)

meaning of global ensemble averages in order to identify realistic and artificial results. In the next sections, we will consider these two different options. From the basic technical point of view, they are actually more or less equivalent (Delle Site and Praprotnik 2017). As a final general remark, it must be underlined that AdResS, in any of its variations, turned out to be conceptually and computationally far more robust than other similar methods (Ensing et al. 2007; Heyden et al. 2007; Mones et al. 2015). In fact, beyond the basic coupling structure, several additional technical aspects were also implemented and tested; relevant examples are the design of a diffusing high-resolution region where the high resolution follows a solvated molecule in space (Praprotnik et al. 2007a) or the design of high-resolution regions of adjustable size (Kreis et al. 2016b), the use of the method for the calculation of potential of mean force for the solvation of relevant biomolecules (Fiorentini et al. 2017) or for the calculation of the chemical potential of liquids and mixtures (Agarwal et al. 2014).

## 3  Grand Canonical AdResS: GC-AdResS

Further methodological developments of the initial idea based on force interpolation led to the definition of a space-dependent thermodynamic force acting on to the centers of mass of molecules in the hybrid region. Such a force imposes a priori a proper thermodynamic equilibrium and a correct exchange of particles between the AT and CG region (Poblete et al. 2010; Fritsch et al. 2012a). Later on, it has been shown that the thermodynamic force and the work of the thermostat exactly

give the proper chemical potential of coupling even when coarse-grained models reproduce only density and temperature that is only macroscopic thermodynamics. Thus, AdResS was reframed within a grand canonical-like approach (Wang et al. 2012, 2013). Finally, the coupling of the atomistic region to the rest of the system has been rationalized in such a way that a rigorous formalization of a control error is employed to check whether or not the textbook definition of grand canonical system (see, e.g., the procedure in Huang 1987) is satisfied (GC-AdResS). The corresponding formula for the control of the error opened the gate to the conceptual embedding of GC-AdResS into the so-called Bergman-Lebowitz model of open systems (Lebowitz and Bergmann 1957; Bergmann and Lebowitz 1955) and hence to the definition of a Liouville equation for a system that exchanges particles with a reservoir. This led to the physically correct definition of ensemble average time correlation functions for open systems (Agarwal et al. 2015; Agarwal and Delle Site 2015, 2016). In this context, a surprising result of an unexpected locality in the IR spectra of water was found (Agarwal and Delle Site 2016), which is due to the interplay between local quantum effect and global structural properties for the solvation of hydrophobic molecules (Agarwal et al. 2017). Another success of GC-AdResS is its application to ionic liquids, due to the highly challenging role of electrostatics (Krekeler and Delle Site 2017). We could show that independently from the fact whether the coarse-grained model is charged or not, the AT region reproduces always the results of an equivalent subregion of a large full atomistic simulation. The absence of any atomistic degree of freedom outside the atomistic region allowed us to confirm and quantify the structure of the cage within which an ion rattles (Jabes et al. 2018). Here, it should be underlined that the fact that the method applied is force-based plays no more a role for the conceptual justification of the method, since the embedding of the algorithm into the grand canonical frame, and its correspondence to the Bergman-Lebowitz model of open system applies to any possible form of coupling between the AT region and the reservoir. Along such lines, recent theoretical progress has led to the construction of a numerical protocol for allowing the adaptive resolution simulation of molecules with electrons where physical consistency at statistical mechanics level is strictly assured by exact formulas of control of the approximations (Delle Site 2018). In general, AdResS can handle systems in equilibrium and beyond, and, in the next sections, the idea of system with open boundaries is moved even forward to a truly open systems such as coupling to continuum or an ideal gas.

## 4  AdResS and Its Coupling to Continuum: Open Boundary MD (OBMD)

Another extension of the force-based AdResS that allows for simulations of open molecular systems that exchange mass, momentum, and energy with their surroundings has been coined open boundary MD (OBMD) (Delgado-Buscalioni et al. 2015; Sablić et al. 2016, 2017a, b). The OBMD methodology, which is an offspring of two linear-momentum-preserving methods, i.e., open MD (Flekkoy et al. 2005;

Delgado-Buscalioni 2012) and AdResS, enables us to perform equilibrium MD simulations in the grand canonical ensemble as well as non-equilibrium fluid flow simulations. The flow is introduced via an external boundary condition, while the equations of motion for the bulk remain unaltered.

In the OBMD simulations, the MD domain is typically sandwiched in between two (auxiliary) buffer domains. The latter two act as mass reservoirs and hence enable the former to exchange particles, momentum, and energy with its surrounding. Thus, the OBMD system is opened, i.e., not periodic, in the coupling direction. Molecules freely move between the MD and buffer domains. Moreover, in the buffers, the molecules change their resolution from the fine- (close to the MD domain) to the coarse-grained resolution (at the outer boundaries of the simulation box). The underlying idea of the resolution change is that AdResS allows for the insertion of molecules of arbitrary size into the system. In the coarse-grained parts of the buffers, relatively large molecules can thus be easily inserted due to soft effective interactions among coarse-grained beads (Delgado-Buscalioni and Coveney 2003; De Fabritiis et al. 2004). Then, as the molecules move toward the MD domain, they gain the fine-grained details owing to AdResS used in the buffers. Molecules are deleted once they leave the outer boundary of a given buffer, and new molecules are inserted to achieve the mass balance, i.e., to have a desired average density in the system.

The OBMD setup can be further extended to computational fluid dynamics (CFD) description of a liquid, i.e., Navier-Stokes (NS) are used to describe the dynamics of the liquid on the continuum scale (going beyond the particle-based modeling of the liquid) (Delgado-Buscalioni et al. 2008, 2009). Such hybrid MD/CFD approaches are especially useful for simulations of the transport of nanoparticles through fluids. Simulations using MD can capture the atomistic details of the nanoparticle-liquid interface, but due to their computational cost, they cannot be extended, in the foreseeable future, to the macroscale regime of the full flow field. In turn, continuum descriptions, using the NS equations may capture the macroscale behavior of the flow, but they fail to represent accurately the flow field at the nanoparticle surface. The hybrid approaches, on the other hand, combine the powerful features of both descriptions, i.e., the ability to describe the macroscale behavior of the flow as well as accurate boundary conditions around nanoparticles (Walther et al. 2012).

While technically and (when considering the AT region as a grand ensemble) conceptually equivalent to the force-based approach, the energy-based approach (H-AdResS), presented in the next section, becomes particularly relevant if one wants to perform MC adaptive resolution simulations.

## 5　Hamiltonian-Based AdResS: H-AdResS

The AdResS method ensures that Newton's 3rd law is exactly satisfied everywhere within the simulation box. As a consequence, it is impossible to interpolate the potential energy using a position-dependent switching function and simultaneously

preserve Newton's 3rd Law. This, as shown in previous sections, proves not to be a major practical problem (Wang et al. 2013). However, without a global Hamiltonian, it is unfeasible to technically carry out NVE or MC simulations in a straightforward way. In the Hamiltonian AdResS (H-AdResS) method (Potestio et al. 2013a, b; Español et al. 2015), the constraint of global momentum conservation of momentum is relaxed in order to overcome the limitations of AdResS reported above. In H-AdResS, total energy of each molecule is expressed in terms of a position-dependent function:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \left\{ \lambda_\alpha V_\alpha^{AT} + (1 - \lambda_\alpha) V_\alpha^{CG} \right\} \tag{2}$$

where $\mathcal{K}$ is the (all-atom) kinetic energy of the molecules, $V^{int}$ is the interaction internal to the molecules, and

$$\begin{cases} V_\alpha^{AT} \equiv \dfrac{1}{2} \sum_{\beta, \beta \neq \alpha}^{N} \sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|) \\ V_\alpha^{CG} \equiv \dfrac{1}{2} \sum_{\beta, \beta \neq \alpha}^{N} V^{CG}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|) \\ \lambda_\alpha = \lambda(\mathbf{R}_\alpha) \end{cases}$$

As in the AdResS scheme, the switching function $\lambda$ goes from 0 (purely CG) to 1 (purely AT). The force acting on atom $i$ in molecule $\alpha$ is obtained through differentiation of the Hamiltonian in Eq. 2. In particular, differently from standard AdResS in this case, one has a further term: $\left[ V_\alpha^{AT} - V_\alpha^{CG} \right] \nabla_{\alpha i} \lambda_\alpha$. This term is asymmetric and locally breaks Newton's 3rd law. This force, which is nonzero only in the hybrid region, is proportional to the difference between the potential energies of a molecule in the AT and the CG representation. The effect of this term is to push molecules into one of the two bulk regions, similar to AdResS, where the thermodynamic force has been introduced. Formally, the solution to this is to introduce a compensation term in such a way that

$$\left. \frac{d \Delta H(\lambda)}{d\lambda} \right|_{\lambda = \lambda_\alpha} = \left\langle \left[ V_\alpha^{AT} - V_\alpha^{CG} \right] \right\rangle_{\mathbf{R}_\alpha} \tag{3}$$

The subscript in the average indicates that the latter has to be performed constraining the CG site of molecule $\alpha$ in the position $\mathbf{R}_\alpha$. This correction leads to a position-dependent correction term in the hybrid zone:

$$\left\langle \left[ V_\alpha^{AT} - V_\alpha^{CG} \right] \right\rangle_{\mathbf{R}_\alpha} \simeq \frac{1}{N} \left\langle \left[ V^{AT} - V^{CG} \right] \right\rangle_{\lambda'} \tag{4}$$

where $\lambda' \equiv \lambda(\mathbf{R}_\alpha)$ is the same for all molecules.

Formally, this can be seen as a position-dependent compensation term to the change in the *Helmholtz free energy* difference between the AT and the CG system (Kirkwood 1935). Therefore, it is possible to calculate the compensating function needed to restore, on average, Newton's 3rd law by performing a Kirkwood thermodynamic integration and add it to the Hamiltonian, so that an effective global Hamiltonian for MD (Potestio et al. 2013a) and MC (Potestio et al. 2013b) approaches, resulting in

$$\Delta H(\lambda) \equiv \Delta\mu(\lambda) = \frac{\Delta F(\lambda)}{N} + \frac{\Delta p(\lambda)}{\rho^{\star}} \tag{5}$$

Practically, while the free energy difference between the AT and the CG description is correctly reproduced, details of the local shape of the compensations force are determined by a few iterations. This ensures perfectly flat density profiles. In general, for such methods global ensemble averages must be carefully considered, since the transition region always represents an artificial filter. Like for AdResS this has been done in very detail for H-AdResS in a recent publication (Español et al. 2015). The general Bergman and Lebowitz framework allows to include H-AdResS within an overall adaptive resolution framework. As for (GC-)AdResS (Poma and Delle Site 2010, 2011; Potestio and Delle Site 2012; Agarwal and Delle Site 2015, 2016; Agarwal et al. 2017) also H-AdResS can be employed to put the coarse-grained resolution in connection with a quantum path integral description of molecules (Kreis et al. 2016a, 2017). The technical advantage in this case is that the presence of a global Hamiltonian allows for a straightforward quantization of the Hamiltonian in the high-resolution region. Other examples are the appropriate treatment of electrostatic interactions in the AT and CG region (Heidari et al. 2016) or the connection to relative entropy coarse-graining methodologies (Kreis and Potestio 2016).

## 6    AdResS, GC-AdResS, and H-AdResS: Examples of Complementary Approaches in a Unified Framework

All AdResS versions are equivalent when considering the AT region as an open system embedded in a reservoir (Agarwal et al. 2015) and that they are equivalent regarding the technical efficiency. Moreover, recent works (Kreis et al. 2014; Zhu et al. 2016) show that a formal connection can be drawn between force-based methods for dual-resolution simulations, such as AdResS, and energy-based methods, such as H-AdResS. Specifically, in Kreis et al. (2014), it was shown that the force-based version of the H-AdResS scheme, i.e., without the drift force, has the same statistical properties of a H-AdResS simulation with a colored noise. Furthermore, mathematical analysis of the asymptotic properties of the force-based and Hamiltonian-based versions (Zhu et al. 2016) reveals that they are equivalent under the condition of large coarse-grained region compared to the atomistic region and (at the same time) large atomistic region compared to the hybrid region. Such

connections prove the mathematical and physical solidity of the AdResS idea in general, which accompanied to its computational efficiency makes the method a powerful tool of simulation and analysis to address the modern challenges in different fields of applied molecular science, as illustrated in the next section.

The AdResS development has meanwhile reached a mature stage at which we can readily apply the methodology to tackle interesting problems in life and material sciences. For instance, AdResS has been applied to study biomolecular systems such as solvated proteins (Fogarty et al. 2015; Kreis et al. 2016b; Fogarty et al. 2016; Tarenzi et al. 2017) and DNA molecules (Zavadlav et al. 2015a, 2017a; Netz et al. 2016). In those examples, we have used a 1-to-1 molecular mapping, in which one coarse-grained bead corresponds to one atomistic water molecule. The next stage of development has involved coupling of atomistic to supramolecular water models such as the widely used MARTINI force-field (Marrink et al. 2004; Marrink et al. 2007; Marrink and Tieleman 2013), where four water molecules are represented with one coarse-grained bead (4-to-1 molecular mapping). Here, water changes its resolution from four molecules to one coarse-grained particle and vice versa adaptively on-the-fly (Zavadlav et al. 2014a, b, 2015b, 2016a; Nagarajan et al. 2013); see Fig. 2. To this end, we have developed a dynamic clustering algorithm SWINGER that can concurrently assemble, disassemble, and reassemble water bundles, consisting of several water molecules (Zavadlav et al. 2016b). Thus, it allows for a seamless coupling between any standard atomistic and supramolecular water models, e.g., dissipative particle dynamics (DPD) (Hoogerbrugge and Koelman 1992; Español and Warren 1995), in adaptive resolution simulations (Zavadlav and Praprotnik 2017). This multiscale approach paves the way for efficient multiscale simulations of biomolecular systems without compromising the accuracy of atomistic water models (Zavadlav et al. 2017b).
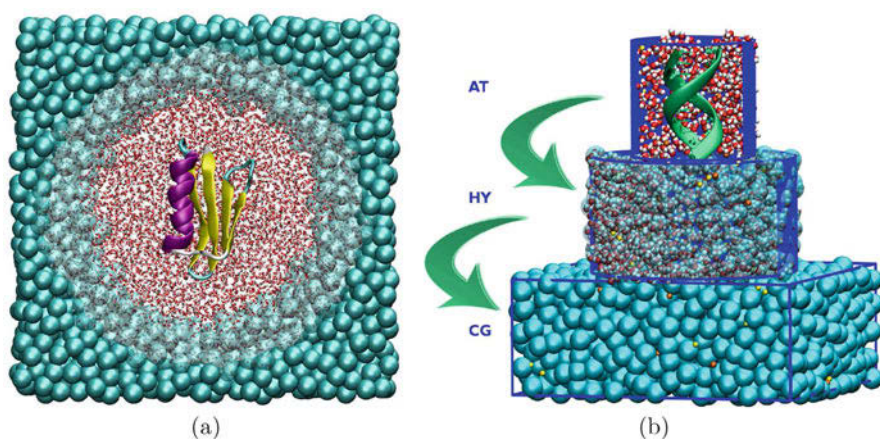


**Fig. 2**  (**a**) AdResS simulation of an atomistic protein in MARTINI water. (**b**) AdResS simulation of an atomistic DNA molecule in MARTINI water. (Reprinted from Zavadlav et al. 2014a, 2016a)

A recent use of AdResS, in its GC-AdResS version, as a tool to investigate the structural locality of ionic liquids, has allowed to unambiguously conclude that specific atom-atom electrostatic interactions do not play a role in the formation of structures on spatial scales beyond the ion-ion distance. Such a conclusion is rather surprising because in current literature, it is actually postulated the opposite, although purely coarse-grained simulations (without explicit charges) had already put the postulate in discussion (Krekeler and Delle Site 2017; Jabes et al. 2018). The study allowed to draw a picture of the essential feature of the liquid structure for a relevant class of ionic liquids (see Fig. 3). Further on, a particularly promising direction to take full advantage of the coupling between models at different resolution is to employ in the CG region the most computationally inexpensive model available, relying on the fact that the compensating external field (thermodynamic force in AdResS, free energy compensation (FEC) in H-AdResS) guarantees that the AT region will reproduce the correct thermodynamics. The extreme case of a model with few, simple interactions is the ideal gas, to be interpreted as a collection of thermostatted noninteracting particles. In Kreis et al. (2015), this possibility was explored by simulating a water system where 6526 water molecules were modeled at the atomistic level in the high-resolution region, while in the CG region, they were subject to no interaction other than the Langevin thermostat. The results show that neither the structure nor the dynamics of the fluid in the AT region of the water-ideal gas hybrid system are compromised. This can be seen from the fully atomistic pair correlation functions and diffusion profiles of the H-AdResS simulation as reported in Fig. 4. These RDFs and diffusion profiles are fully consistent with those measured in completely atomistic reference simulations. Another application which goes beyond the previous examples is the triple-scale simulation of an enzyme in solution with a binding ligand (Fogarty et al. 2016). There the center part of the enzyme, the ligand, and the surrounding water are treated on an all-atom level, while
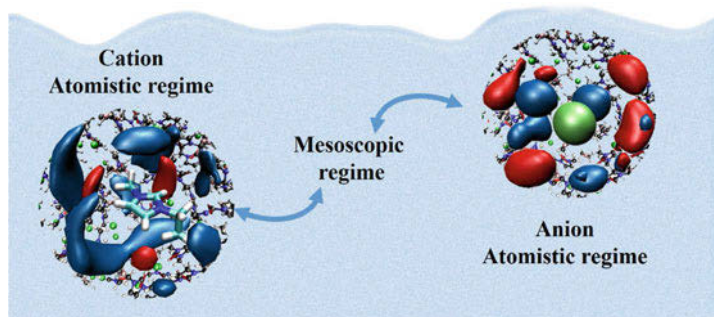


**Fig. 3** Cations and anions in imidazolium-based ionic liquids need only a first shell of neighboring ions to build typical structures of the liquid. Such atomistic islands are embedded in a thermodynamic bath without the need of explicit electrostatic interactions. This picture deviates from the intuitive argument put forward in many papers in the field that explicit electrostatic interactions play a key role in the formation of typical liquid structures, even at local level. (This figure is an adaptation from Jabes et al. 2018)
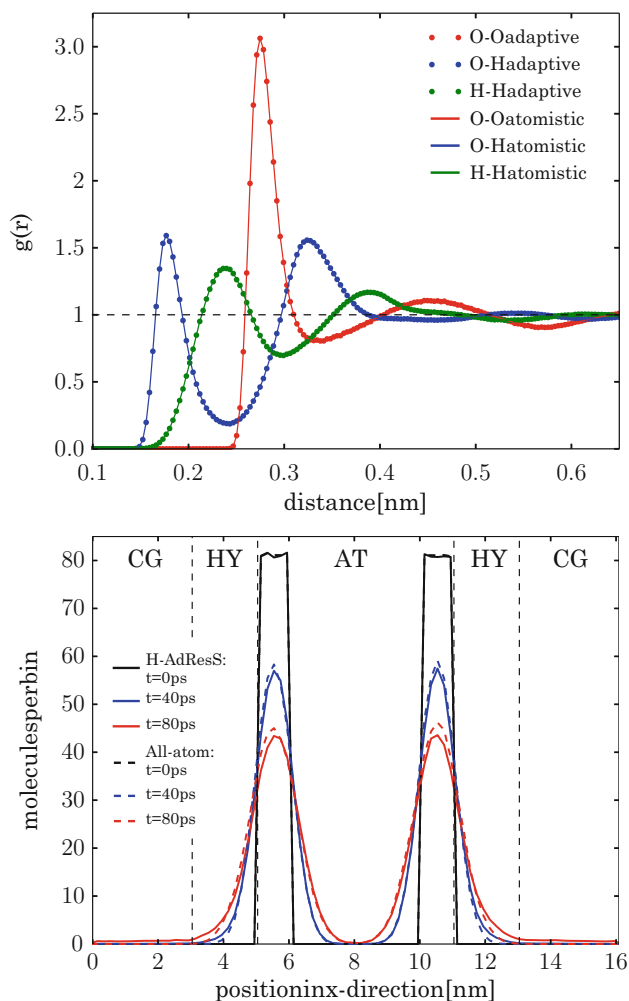
**Fig. 4** Top: RDFs for pure water and for the atomistic region of the H-AdResS water-ideal gas with Kirkwood TI-based Gibbs FEC. Since a rigid water model is employed, all RDFs consider only intermolecular atom pairs. Bottom: diffusion profiles in H-AdResS simulations of the water-ideal gas system and in fully atomistic reference simulations of SPC/E water: time evolution of the position of molecules initially located in a 1-nm-wide slab in the atomistic region, immediately adjacent to the HY region. The y-axis is the absolute number of these molecules whose center of mass $X$ coordinate is in a given bin at the given time. (Figures from Kreis et al. 2015)

the water molecules further away are treated on a coarse-grained level and the rest of the residues of the enzyme is considered on an elastic network level. Finally, on the practical side, it must be reported that all the various method start to be included in some of the most popular simulation codes used within the community. In particular, AdResS, H-AdResS, and OBMD are implemented in ESPResSo++

(Halverson et al. 2013; Guzman et al. 2017), AdResS, H-AdResS, and GC-AdResS are implemented in GROMACS (Abraham et al. 2015; Fritsch et al. 2012a), and H-AdResS is implemented also in LAMMPS (Plimpton 1995; Heidari et al. 2016).

# 7 Perspectives

By now, the general AdResS concept with its variants is well established and has been shown to be robustly applicable to a variety of important problems. Beyond that new developments, which require more or less effort of method development or code adjustment open the path to new applications and opportunities. In fact, the possibility of treating molecules with electrons would allow AdResS to become a tool that can span all the relevant scales of molecular science up to the fluid dynamics. In this perspective, H-AdResS has been connected to QM/MM approaches in which a small sub-domain of the system of interest is treated at the ab initio level, while the rest is described with a classical potential (Boereboom et al. 2016). Furthermore, as anticipated in the section dedicated to GC-AdResS, it has been proposed to go even beyond that and use AdResS (in any form GC-AdResS or H-AdResS) as a physical rigorous platform for a truly grand canonical method for molecules with electrons (Delle Site 2018). A resulting electronic quantum mechanical AdResS approach (el-QM-AdResS) would allow to treat local quantum events, such as conformational deformations of atomistic structures, chemical reactions, specific solvation, and atom-atom bonding effects, to name a few, while assuring the correct statistical mechanics electronic structure and macroscopic thermodynamics. The coupling of any region of resolution to an ideal gas allows the easy treatment of systems, which are driven by an external supply of matter, such as controlled growth or structure formation.

# References

Abraham MJ, Murtola T, Schulz R, Pall S, Smith JC, Hess B, Lindahl E (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25

Agarwal A, Delle Site L (2015) Path integral molecular dynamics within the grand canonical-like adaptive resolution technique: simulation of liquid water. J Chem Phys 143:094102

Agarwal A, Delle Site L (2016) Grand-canonical adaptive resolution centroid molecular dynamics: implementation and application. Comp Phys Commun 206:26

Agarwal A, Wang H, Schütte C, Delle Site L (2014) Chemical potential of liquids and mixtures via adaptive resolution simulation. J Chem Phys 141:034102

Agarwal A, Zhu J, Hartmann C, Wang H, Delle Site L (2015) Molecular dynamics in a grand ensemble: Bergmann-Lebowitz model and adaptive resolution simulation. New J Phys 17:083042

Agarwal A, Clementi C, Delle Site L (2017) Path integral-GC-AdResS simulation of a large hydrophobic solute in water: a tool to investigate the interplay between local microscopic structures and quantum delocalization of atoms in space. Phys Chem Chem Phys 19:13030–13037

Bergmann PG, Lebowitz JL (1955) New approach to nonequilibrium processes. Phys Rev 99:578

Boereboom JM, Potestio R, Donadio D, Bulo RE (2016) Toward hamiltonian adaptive QM/MM: accurate solvent structures using many-body potentials. J Chem Theory Comput 12:3441–3448

De Fabritiis G, Delgado-Buscalioni R, Coveney PV (2004) Energy controlled insertion of polar molecules in dense fluids. J Chem Phys 121:12139

Delgado-Buscalioni R (2012) Tools for multiscale simulation of liquids using open molecular dynamics. Numerical analysis of multiscale computations. Springer, Berlin/Heidelberg, pp 145–166

Delgado-Buscalioni R, Coveney PV (2003) Usher: an algorithm for particle insertion in dense fluids. J Chem Phys 119:978–987

Delgado-Buscalioni R, Kremer K, Praprotnik M (2008) Concurrent triple-scale simulation of molecular liquids. J Chem Phys 128:114110

Delgado-Buscalioni R, Kremer K, Praprotnik M (2009) Coupling atomistic and continuum hydrodynamics through a mesoscopic model: application to liquid water. J Chem Phys 131:244107

Delgado-Buscalioni R, Sablić J, Praprotnik M (2015) Open boundary molecular dynamics. Eur Phys J Spec Top 224:2331–2349

Delle Site L (2007) Some fundamental problems for an energy-conserving adaptive resolution molecular dynamics scheme. Phys Rev E 76:047701

Delle Site L (2018) Grand canonical adaptive resolution simulation for molecules with electrons: a theoretical framework based on physical consistency. Comp Phys Commun 222:94–101

Delle Site L, Praprotnik M (2017) Molecular systems with open boundaries: theory and simulation. Phys Rep 693:1–56

Ensing B, Nielsen SO, Moore PB, Klein ML, Parrinello M (2007) Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics. J Chem Theory Comput 3:1100

Español P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. Europhys Lett 30:191

Español P, Delgado-Buscalioni R, Everaers R, Potestio R, Donadio D, Kremer K (2015) Statistical mechanics of hamiltonian adaptive resolution simulations. J Chem Phys 142:064115

Fiorentini R, Kremer K, Potestio R, Fogarty AC (2017) Using force-based adaptive resolution simulations to calculate solvation free energies of amino acid sidechain analogues. J Chem Phys 146:244113

Flekkoy EG, Delgado-Buscalioni R, Coveney PV (2005) Flux boundary conditions in particle simulations. Phys Rev E 72:026703

Fogarty AC, Potestio R, Kremer K (2015) Adaptive resolution simulation of a biomolecule and its hydration shell: structural and dynamical properties. J Chem Phys 142:195101

Fogarty AC, Potestio R, Kremer K (2016) A multi-resolution model to capture both global fluctuations of an enzyme and molecular recognition in the ligand-binding site. Proteins 84:1902–1913

Fritsch S, Junghans C, Kremer K (2012a) Structure formation of toluene around c60: implementation of the adaptive resolution scheme (adress) into gromacs. J Chem Theory Comput 8:398–403

Fritsch S, Poblete S, Junghans C, Ciccotti G, Delle Site L, Kremer K (2012b) Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir. Phys Rev Lett 108:170602

Guzman HV, Junghans C, Kremer K, Stuehn T (2017) Scalable and fast heterogeneous molecular simulation with predictive parallelization schemes. Phys Rev E 96:053311

Halverson JD, Brandes T, Lenz O, Arnold A, Bevc S, Starchenko V, Kremer K, Stuehn T, Reith D (2013) Espresso++: a modern multiscale simulation package for soft matter systems. Comput Phys Commun 184:1129–1149

Heidari M, Cortes-Huerto R, Donadio D, Potestio R (2016) Accurate and general treatment of electrostatic interaction in hamiltonian adaptive resolution simulations. Eur Phys J Spec Top 225:1505–1526

Heyden A, Lin H, Truhlar DG (2007) Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. J Phys Chem B 111:2231–2241

Hoogerbrugge PJ, Koelman JMVA (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhys Lett 19:155–160

Huang K (1987) Statistical mechanics. Wiley, New York

Jabes BS, Krekeler C, Klein R, Delle Site L (2018) Probing spatial locality in ionic liquids with the grand canonical adaptive resolution molecular dynamics technique. J Chem Phys 148:193804

Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3(5):300

Kreis K, Potestio R (2016) The relative entropy is fundamental to adaptive resolution simulations. J Chem Phys 145:044104

Kreis K, Donadio D, Kremer K, Potestio R (2014) A unified framework for force-based and energy-based adaptive resolution simulations. EPL 108(3):30007

Kreis K, Fogarty AC, Kremer K, Potestio R (2015) Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations. Eur Phys J Spec Top 224:2289–2304

Kreis K, Potestio R, Kremer K, Fogarty AC (2016a) Adaptive resolution simulations with self-adjusting high-resolution regions. J Chem Theory Comput 12:4067–4081

Kreis K, Tuckerman ME, Donadio D, Kremer K, Potestio R (2016b) From classical to quantum and back: a hamiltonian scheme for adaptive multiresolution classical/path-integral simulations. J Chem Theory Comput 12:3030–3039

Kreis K, Kremer K, Potestio R, Tuckerman ME (2017) From classical to quantum and back: hamiltonian adaptive resolution path integral, ring polymer, and centroid molecular dynamics. J Chem Phys 147:244104

Krekeler C, Delle Site L (2017) Towards open boundary molecular dynamics simulation of ionic liquids. Phys Chem Chem Phys 19:4701–4709

Lebowitz JL, Bergmann PG (1957) Irreversible Gibbsian ensembles. Ann Phys 1:1

Marrink SJ, Tieleman DP (2013) Perspective on the MARTINI model. Chem Soc Rev 42:6801–6822

Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108:750–760

Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111:7812–7824

Mones L, Jones A, Götz AW, Laino T, Walker RC, Leimkuhler B, Csany G, Bernstein N (2015) The adaptive buffered force QM/MM method in the CP2K and AMBER software packages. J Comp Chem 36:633

Nagarajan A, Junghans C, Matysiak S (2013) Multiscale simulation of liquid water using a four-to-one mapping for coarse-graining. J Chem Theory Comput 9:5168–5175

Netz PA, Potestio R, Kremer K (2016) Adaptive resolution simulation of oligonucleotides. J Chem Phys 145:234101

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comp Phys 117:1

Poblete S, Praprotnik M, Kremer K, Delle Site L (2010) Coupling different levels of esolution in molecular simulations. J Chem Phys 132:114101

Poma A, Delle Site L (2010) Classical to path-integral adaptive resolution in molecular simulation: towards a smooth quantum-classical coupling. Phys Rev Lett 104:250201

Poma AB, Delle Site L (2011) Adaptive resolution simulation of liquid parahydrogen: testing the robustness of the quantum-classical adaptive coupling. Phys Chem Chem Phys 13:10510

Potestio R, Delle Site L (2012) Quantum locality and equilibrium properties in low-temperature parahydrogen: a multiscale simulation study. J Chem Phys 136:054101

Potestio R, Español P, Delgado-Buscalioni R, Everaers R, Kremer K, Donadio D (2013a) Monte carlo adaptive resolution simulation of multicomponent molecular liquids. Phys Rev Lett 111:060601

Potestio R, Fritsch S, Español P, Delgado-Buscalioni R, Kremer K, Everaers R, Donadio D (2013b) Hamiltonian adaptive resolution simulation for molecular liquids. Phys Rev Lett 110:108301

Praprotnik M, Delle Site L, Kremer K (2005) Adaptive resolution molecular- dynamics simulation: changing the degrees of freedom on the fly. J Chem Phys 123:224106

Praprotnik M, Delle Site L, Kremer K (2006) Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. Phys Rev E 73:066701

Praprotnik M, Matysiak S, Delle Site L, Kremer K, Clementi C (2007a) Adaptive resolution simulation of liquid water. J Phys Condens Matter 19:292201

Praprotnik M, Delle Site L, Kremer K (2007b) A macromolecule in a solvent: adaptive resolution molecular dynamics simulation. J Chem Phys 126:134902

Praprotnik M, Delle Site L, Kremer K (2008) Multiscale simulation of soft matter: from scale bridging to adaptive resolution. Ann Rev Phys Chem 59:545–571

Sablić J, Praprotnik M, Delgado-Buscalioni R (2016) Open boundary molecular dynamics of sheared star-polymer melts. Soft Matter 12:2416–2439

Sablić J, Delgado-Buscalioni R, Praprotnik M (2017a) Application of the eckart frame to soft matter: rotation of star polymers under shear flow. Soft Matter 13:6988–7000

Sablić J, Praprotnik M, Delgado-Buscalioni R (2017b) Deciphering the dynamics of star molecules in shear flow. Soft Matter 13:4971–4987

Tarenzi T, Calandrini V, Potestio R, Giorgetti A, Carloni P (2017) Open boundary simulations of proteins and their hydration shells by hamiltonian adaptive resolution scheme. J Chem Theory Comput 13:5647–5657

Walther JH, Praprotnik M, Kotsalis EM, Koumoutsakos P (2012) Multiscale simulation of water flow past a C540 fullerene. J Comput Phys 231:2677–2681

Wang H, Hartmann C, Schütte C, Delle Site L (2013) Grand-canonical-Like molecular-dynamics simulations by using an adaptive-resolution technique. Phys Rev X 3:011018

Wang H, Schütte C, Delle Site L (2012) Adaptive resolution simulation (AdResS): a smooth thermodynamic and structural transition from atomistic to coarse grained resolution and vice versa in a grand canonical fashion. J Chem Theory Comput 8:2878–2887

Zavadlav J, Praprotnik M (2017) Adaptive resolution simulations coupling atomistic water to dissipative particle dynamics. J Chem Phys 147:114110

Zavadlav J, Melo MN, Cunha AV, de Vries AH, Marrink SJ, Praprotnik M (2014a) Adaptive resolution simulation of MARTINI solvents. J Chem Theory Comput 10:2591–2598

Zavadlav J, Melo MN, Marrink SJ, Praprotnik M (2014b) Adaptive resolution simulation of an atomistic protein in MARTINI water. J Chem Phys 140:054114

Zavadlav J, Melo MN, Marrink SJ, Praprotnik M (2015a) Adaptive resolution simulation of polarizable supramolecular coarse-grained water models. J Chem Phys 142:244118

Zavadlav J, Podgornik R, Praprotnik M (2015b) Adaptive resolution simulation of a DNA molecule in salt solution. J Chem Theory Comput 11:5035–5044

Zavadlav J, Marrink SJ, Praprotnik M (2016a) Adaptive resolution simulation of supramolecular water: the concurrent making, breaking, and remaking of water bundles. J Chem Theory Comput 12:4138–4145

Zavadlav J, Podgornik R, Melo MN, Marrink SJ, Praprotnik M (2016b) Adaptive resolution simulation of an atomistic DNA molecule in MARTINI salt solution. Eur Phys J Spec Top 225:1595–1607

Zavadlav J, Bevc S, Praprotnik M (2017a) Adaptive resolution simulations of biomolecular systems. Eur Biophys J 46:821–835

Zavadlav J, Podgornik R, Praprotnik M (2017b) Order and interactions in DNA arrays: multiscale molecular dynamics simulation. Sci Rep 7:4775–4786

Zhu J, Klein R, Delle Site L (2016) Adaptive molecular resolution approach in hamiltonian form: n asymptotic analysis. Phys Rev E 94:043321

# Data-Driven Methods in Multiscale Modeling of Soft Matter

# 67

Tristan Bereau

## Contents

### Abstract

As in many other scientific fields, data-driven methods are rapidly impacting multiscale modeling. This chapter will illustrate some of the many ways advanced statistical models and a data-centric perspective help augmenting computer simulations in soft matter. A specific focus on force fields, sampling, and simulation analysis is presented, taking advantage of machine learning, high-throughput schemes, and Bayesian inference.

## 1　Introduction

Advanced statistical models are rapidly impregnating many technological and scientific fields, from the automobile industry to robotics to particle physics. Not only do novel data-driven methods offer new perspectives on approaching long-

T. Bereau (✉)
Theory Group, Max Planck Institute for Polymer Research, Mainz, Germany
e-mail: bereau@mpip-mainz.mpg.de

standing problems, they hold the promise of accelerating the pace of research. Materials science is one such field, where data is likely to accelerate computational rational design. The decisive impact of materials design in various aspects of our society has led to large-scale strategies – among others the Materials Genome Initiative (Jain et al. 2013). These recent efforts are already bearing fruit in various disciplines of hard condensed matter, inorganic chemistry, and also semiconductor physics (Curtarolo et al. 2013). Interestingly, little has happened in soft matter.

The slow development of computational materials design in soft matter likely precisely arises from what makes these systems unique: the prominent role of thermal fluctuations. Soft matter systems display weak characteristic energies on par with thermal energy, $k_B T$, leading to fascinating phenomena, such as self-assembly. On the other hand, thermal fluctuations obscure the link between the chemistry and materials properties, because of the complex interplay of a system with its environment. This makes computational materials discovery for soft matter all the more challenging (Bereau et al. 2016).

Modeling soft matter systems is traditionally approached using multiscale simulations. They bridge the relevant length and time scales of the system: from quantum, to classical atomistic, to coarse-grained (CG), and to continuum resolutions. These methods are all entrenched within certain physical laws and symmetries. They stand at odds with purely data-driven methods, which typically contain little physics a priori but are instead mostly empirical. Can we benefit by combining these two paradigms?

This chapter discusses recent examples that apply data-driven methods to *augment* multiscale modeling in soft matter. Here, I will emphasize how advanced statistical models can help improve existing methodologies or offer new perspectives. The chapter describes efforts in building better force fields, tackling sampling challenges, but also efficiently analyzing computer simulations. In each case, significant progress is achieved by a variety of methods, such as machine learning (ML), high-throughput schemes, and Bayesian inference. This chapter will assume prior exposure to computer simulations – it is intended to help the simulator better grasp the benefits of introducing data-driven methods in their research.

## 2 Force Fields

Force fields lie at the heart of classical particle-based modeling. When numerically integrating Newton's equations of motion, the force field dictates how particles interact over time. As such, the force field encodes all the physics and chemistry of the model, no less. Accuracy here is critical because it determines the aggregate behavior of the system after heaps of integration steps. Emergent complexity arises from countless evaluations of $\mathbf{F} = m\mathbf{a}$. In this sense, the force field links the system's chemical composition to its long-time properties, such as free energies or kinetic properties. The corollary to this critical role is the attention force fields have received in the last three to four decades (Maple et al. 1988; Halgren 1992; Halgren and Damm 2001; Wang et al. 2001; Ponder and Case 2003; Mackerell 2004).

Force fields map a particle configuration to interaction energies and forces, leading to the coveted *potential energy surface*. The mapping ought to hit an appropriate balance between accuracy and computational investment: the physics should be described appropriately at small numerical cost. For instance, a simple spring will capture the limited range of a covalent bond but will evidently fail to describe anharmonic effects. Identifying the sweet spot depends critically on the problem at hand. The other facet of a force field development project entails transferability: given a parametrization among certain configurations, compounds, and environments, to what extent can the resulting model extrapolate to scenarios absent from the training set? In the following, we highlight recent strategies where ML has helped improve force field accuracy and transferability.

## 2.1    Accuracy: Beyond Traditional Basis Sets

Traditionally, most of the functional forms commonly used in molecular mechanics have largely been constrained by computational considerations. Among others, a pairwise decomposition is an appealing treatment of intermolecular interactions but fails to capture some of the many-body physics, as found, for instance, when modeling dispersion (Tkatchenko et al. 2012). Mathematically, this is a basis set problem: the vector space used to construct the force field fails to accurately reproduce all aspects of the underlying potential energy surface.

A striking illustration of the basis set problem arises upon coarse-graining. Coarse-graining reduces the representation of a molecular system by grouping atoms into larger particles or beads. Structure-based coarse-graining aims at a systematic derivation of CG potentials from reference atomistic simulations (Voth 2008; Peter and Kremer 2010; Noid 2013). Several methods exist to derive CG potentials that aim at best reproducing the underlying forces or distribution functions. Examples of these strategies include force matching and iterative Boltzmann inversion. The averaging performed over the degrees of freedom that have been coarse-grained away effectively leads to a potential of mean force (PMF). This PMF is typically a many-body quantity. The many-body aspect holds even when the reference simulation only relies on pairwise interactions, because of correlations owing to the missing degrees of freedom (Rühle et al. 2009). This situation makes the pairwise assumption even more critical in CG models, limiting an accurate description of the structure and thermodynamics.

Unlike standard regression schemes, a machine learning (ML) algorithm does not aim at optimally fitting parameters on a predefined basis set. It instead looks for similarities between training points to interpolate the target property in a high-dimensional feature space. Because the interpolation can always improve with added training points, a specific attribute of ML is its ability to improve its accuracy with added data. We illustrate the concept with kernel ridge regression (Rasmussen and Williams 2006), though neural network-type architectures share a number of aspects (see Behler 2016). Consider the regression of property $p$ of sample $\mathbf{x}$. A kernel machine will consist of the prediction:

$$p(\mathbf{x}) = \sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}), \tag{1}$$

where the sum runs over training samples, $\alpha_i$ is the weight of sample $i$, and $K(\mathbf{x}, \mathbf{x}')$ is the kernel between samples $\mathbf{x}$ and $\mathbf{x}'$. The kernel consists of a similarity measure, or covariance function, between two samples:

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2}\right), \tag{2}$$

where here we chose a Gaussian kernel with Euclidean distance as a measure between two samples. This metric implies that the distance is 1 and 0 for samples that are identical and very different, respectively. The middle part of the equation expresses the kernel as an inner product between the two samples. While the samples are expressed in their input space, also called "representations," the so-called *kernel trick* implicitly maps the samples via $\phi$ into an infinite-dimensional feature space, where the interpolation between samples takes place (Schiilkopf 2001). The optimization of the weights $\boldsymbol{\alpha}$ consists of solving Eq. 1 for the samples in the training set, adding a regularization term $\lambda$:

$$\boldsymbol{\alpha} = (K + \lambda \mathbb{K})^{-1} \mathbf{p}, \tag{3}$$

where $\lambda$ is set by the amount of noise in the reference data. Interestingly, while the basis set limits the possible accuracy of a regression problem, it is largely the relevance of the representation that determines the accuracy of an ML model (Huang and von Lilienfeld 2016).

There is a rapidly growing literature of studies applying machine learning to learning chemical properties, such as atomization energies (Rupp et al. 2012; Ramakrishnan and von Lilienfeld 2017). Adequate training can yield remarkably accurate predictions (Faber et al. 2017). Recent studies have aimed at using ML methods to help optimize a potential energy surface or force field (Li et al. 2017; Huan et al. 2017). When it comes to learning forces, the intrinsic orientation of the vector must be reproduced. Unlike scalar quantities, vectors contain three independent components. Different strategies have been devised to tackle this issue: the derivative of the kernel with respect to particle coordinates (Bartók et al. 2010; Chmiela et al. 2017), local axis systems (Bereau et al. 2015), or covariant kernels (Glielmo et al. 2017).

The use of ML potentials has mostly been applied to replace expensive ab initio MD simulations (Li et al. 2015; Morawietz et al. 2016; Deringer and Csányi 2017), where the computational cost difference between a single-point electronic structure calculation and ML prediction is significant. When aiming at predicting classical reference models however, the gain is smaller. Recent work on coarse-graining two benzene molecules in water indicates a better reproduction of the PMF compared to force matching (John 2016). The cost of the prediction remains significant compared to traditional pairwise potentials, but these results provide potential avenues to break the glass ceiling of pairwise interactions in CG potentials.

## 2.2     Transferability: Across Conformations, Phases, and Compositions

Molecular simulations often exhibit a complex relationship between model parameters and the resulting emergent properties. This obscures the role and impact of force field parameters. For instance, how does the tuning of a Lennard-Jones parameter affect a compound's hydration free energy or a folding timescale? This complex relationship can make force field parametrization a tedious, long, and rather unsystematic process. Systematically understanding the relationship between force field parameter and thermodynamic properties can help automate parametrization methods (Stroet et al. 2017). More often than not, coarser models tend to be more difficult to parametrize, because the missing physics require ad hoc compensations. For instance, most biomolecular atomistic models are additive – they do not explicitly model induction/polarization. Instead, mean-field polarization effects are incorporated effectively by tuning the other force field terms, most importantly Coulomb and Lennard-Jones. This typically comes at the cost of limited phase transferability: not only will they not transfer from the gas to the condensed phase, these models are typically state-point dependent. In other words, they are bound to a limited range of thermodynamic parameters, such as temperature and pressure.

Enhancing the phase transferability of these models is subject to ongoing research – a field where ML can help (Deringer and Csányi 2017) – but not the only strategy. The obscure link between model parameters and emergent properties leads to an unsystematic, largely empirical approach to force field parametrization – the craftsmanship of a biomolecular modeler. As such, developing more automated parametrization schemes offers extremely valuable perspectives: reduced parametrization efforts would speed up and enhance the pace of research in molecular simulations. In the following, two examples from atomistic and coarse-grained modeling illustrate this emerging trend. They both leverage the link between chemical properties and specific force field parameters.

High-resolution models offer a closer, more straightforward link from chemistry to force field parameters. This has motivated the development of force fields from first principles: Electronic structure calculations provide molecular and atomic properties, such as atomic polarizabilities or electrostatic coefficients, used as parameters for classical models (Van Vleet et al. 2016). This framework still requires reference calculations for every new compound considered. One can instead envision relying on the abovementioned use of ML to *predict* these chemical properties. Such a scheme was recently introduced (Bereau et al. 2018) to construct classical intermolecular potentials from atomic polarizabilities, multipole electrostatic coefficients (Bereau et al. 2015), and atomic density parameters. These parameters are fed into a physics-based model based on perturbation theory and an overlap model at long and short ranges, respectively. They lead to a remarkably small number of global parameters that only need tuning across organic compounds once and for all.

Switching to a coarse-grained resolution, the transferable Martini biomolecular force field offers a set of bead types, from which one constructs biomolecules, from proteins to lipids to sugars (Marrink and Tieleman 2013). Charged groups

interact through integer-charge Coulomb interactions. Otherwise, beads interact by means of Lennard-Jones interactions, with a predefined interaction matrix that determines the cross interactions between beads (Marrink et al. 2007). The model aims at capturing the essential thermodynamics of partitioning of chemical groups in different environments. In particular, it relies heavily on the water/octanol partitioning to assign a measure of hydrophobicity to the bead. Though not readily accessible, the water/octanol partitioning can also be predicted: ML models exist to do just that (Tetko et al. 2001). This enables a completely automated parametrization of Martini for small molecules, which both optimizes the mapping from atoms to beads and assigns the most appropriate bead type to every chemical group (Bereau and Kremer 2015). The parametrization scheme was applied to the calculation of solvation free energies for more than 1,000 compounds, clearly illustrating the potential benefits beyond manual parametrizations.

### 2.3    Example: Learning of Atomic Polarizabilities

As an illustrative example, we consider the learning of atomic polarizabilities across small organic molecules, following Bereau et al. (2018). Atomic polarizabilities are estimated using the Hirshfeld ratio, which consists of a spatial integral over the electron density of an atom in a molecule, compared to the corresponding free atom. Reference data consist of quantum chemistry calculations for thousands of isolated small molecules. We refer the interested reader to Bereau et al. (2018) for further technical details. The code to generate the data below can be found in a repository online (Bereau 2018).

   We build an ML model using kernel ridge regression (Eq. 1) and encode atomic environments – the representation – using the Coulomb matrix: a pairwise matrix of inverse distances scaled by the product of atomic numbers (Rupp et al. 2012). The dataset is split between training and test sets to ensure out-of-sample predictions, thereby limiting overfitting. Figure 1a shows a learning curve: the mean absolute error (MAE) as a function of the number of atoms incorporated in the training set. The error systematically decreases with added data. Note the power law behavior. Figure 1b displays the correlation between predicted and reference Hirshfeld ratios for the rightmost ML model shown in panel a. The color-coding distinguishes between chemical elements.

### 3    Sampling

Sampling is the second corner stone of particle-based modeling in soft matter: teasing out a representative subset of conformational space is essential to extract reliable condensed-phase properties, from free energies to kinetics. The difficulty lies in assessing how much sampling is good enough. Umbrella sampling simulations are notoriously challenging, as they often hide slow conformational changes happening on degrees of freedom orthogonal to the reaction coordinate(s) (Neale et al. 2011).
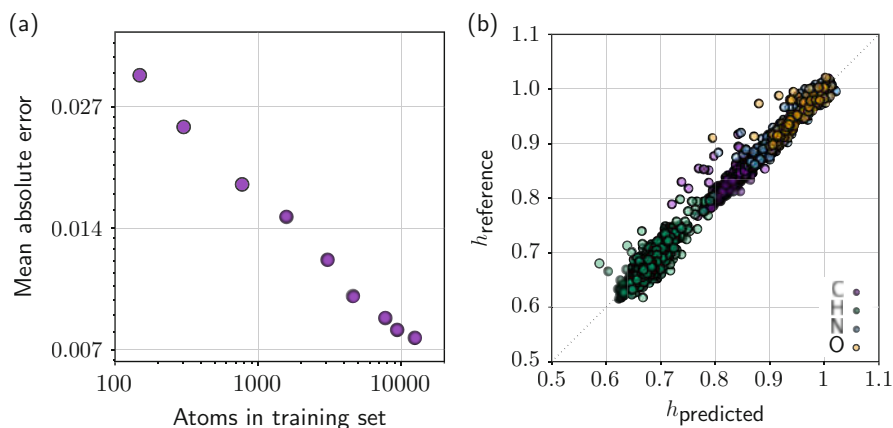
**Fig. 1** (**a**) Saturation curve of the mean absolute error (MAE) as a function of training set size for Hirshfeld ratios. Note that both axes are on a log scale. (**b**) Correlation plot of out-of-sample predictions of the Hirshfeld ratio, $h$, for the ML model with largest training set size. The different chemical elements considered are color-coded

Unsupervised machine learning techniques may help systematically improve the sampling of conformational space within an umbrella sampling protocol (Ferguson et al. 2011).

The last decade has seen a significant leap forward in timescales accessible to computer simulations. Naively, reaching longer timescales means running ever-longer single trajectories. This bodes poorly with high-performance computer clusters, because one cannot parallelize a trajectory in time. A recent paradigm shift consisted in better leveraging the statistics contained in swarms of short trajectories covering the relevant parts of conformational space. For instance, Markov state models (MSMs) discretize the simulation trajectory in conformational space and in time to analyze its long-time kinetics (Noé 2008; Bowman et al. 2013). This framework has been shown to be extremely efficient in leveraging computational resources available – from distributed computing to high-performance clusters – even at a time when dedicated hardware has significantly pushed the state of the art for long trajectories (Shaw et al. 2014). The surge in high-throughput short simulations has helped approach the sampling problem more systematically: an adaptive sampling strategy spawns new simulations from poorly populated regions of conformational space, until convergence is found. Examples include protein-protein interactions (Plattner et al. 2017) and intrinsically disordered proteins (Kukharenko et al. 2016). These ideas rely on a simple concept: it is often easier to locally equilibrate highly diverse seed conformations than waiting for a single trajectory to cross all relevant barriers. This can be extended to a multiscale approach, in which relevant snapshots from computationally efficient CG trajectories are backmapped to provide these seed conformations at the atomistic level. This strategy can help cut down the computational investment of free-energy calculations

by more than tenfold (Menichetti et al. 2017b). Analogously, a more data-driven alternative has been proposed that tries to *extrapolate* possibly interesting new seed conformations from unsupervised machine learning techniques (Chiavazzo et al. 2017).

The discussion so far has focused on sampling conformational space. Recent developments in materials discovery are aiming at exploring chemical compound space – the diversity of chemical compounds – to extract thermodynamic properties. From a simulation perspective, this poses significant challenges due to the compounded issue of sampling both across conformational and chemical compound space. While unattainable at an atomistic resolution for the foreseeable future, coarse-graining can help address this: high-throughput coarse-grained simulations provide an ensemble study of the PMF for the insertion of solute molecules in a lipid bilayer. The study both predicted PMFs for more than 450,000 compounds and identified novel linear relationships between bulk measurements and features of the PMF (Menichetti et al. 2017a).

## 4 Analysis

Everything mentioned so far has focused on improving the quality of computer simulations by improving the force field or the sampling. This section instead consists of extracting insight or information from an *existing* simulation. Advanced data-driven and statistical methodologies have helped develop more robust methods to analyze computer simulations.

Some of the most interesting developments in the analysis of computer simulations have come from approaching the very concept of probability in a new way. The traditional approach to probability theory – the one taught most often at an elementary level – is so-called *frequentist*. It interprets probability from the frequency or propensity of an event to occur. Complementary to this is the Bayesian perspective: how can one infer a reasonable expectation given limited data and/or prior belief? It offers an elegant framework to evaluate the probability of a model $M$, when dealing with limited data $D$, as illustrated by Bayes' theorem:

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}, \tag{4}$$

where $P(M|D)$ and $P(D|M)$ are coined the posterior and the likelihood, respectively. Bayes' theorem has shown extremely useful because while the posterior is typically difficult to evaluate directly, the likelihood is often easier. In addition, it highlights the concept of prior information by means of $P(M)$, which encodes external information we may already hold on the validity of a model. For instance, physical laws and symmetries can naturally be enforced into the prior, effectively biasing the distribution of models to those that satisfy these constraints. Several examples illustrate the conceptual benefits of approaching a problem in a Bayesian framework:

- The weighted histogram method (WHAM) provides a minimum-variance estimator to best estimate the density of states of a system from different simulations (Ferrenberg and Swendsen 1989). These simulations provide complementary information to the system by encompassing a range of temperatures or different values of a collective variable in enhanced sampling. The likelihood incorporates the different Boltzmann distributions, while the prior ensures the normalization of probabilities. A derivation can be found elsewhere (Bereau et al. 2016; Ferguson 2017).
- MSMs build a discrete propagator for the time evolution of a simulation or single-molecule experiment (Noé 2008; Bowman et al. 2013). The simulation trajectory or experimental time series feeds into the likelihood, while the prior incorporates several constraints, most notably detailed balance.
- The MSM of a simulation trajectory can be further tuned to best incorporate external kinetic information. This is useful when a model is known to yield inconsistent kinetics, such as most coarse-grained models. So-called biased MSMs incorporate the coarse reference kinetic information (e.g., folding timescale or mean first passage time) as a prior, thereby selecting more consistent probabilistic models (Rudzinski et al. 2016). This conceptual framework was recently applied to incorporate experimental information to atomistic simulations (Olsson et al. 2017). More generally, the blending of physics-based models with experimental information has recently been subject to increasing interest (Perez et al. 2015, 2017).

Hidden Markov models (HMMs) add to MSMs the possibility of handling unobserved/hidden states (Rabiner and Juang 1986). While these states are not directly visible, the output, which is dependent on the state, is visible. One illustrative analogy consists of a hermit stuck inside a cave: he is attempting to forecast the weather but cannot see the sky outside. His best strategy is then to collect indirect evidence by analyzing the state of a seaweed – probabilistically related to the state of the weather – and thereby to *infer* the hidden state of the weather. HMMs can be thought as a nonlinear filtering process and have been shown to be useful in several studies, from the identification of liquid-ordered and liquid-disordered domains in lipid membrane simulations (Sodt et al. 2014) to the kinetics of protein-protein association (Plattner et al. 2017).

## 5    Outlook

Advanced data-driven methods and data-centric simulation protocols are rapidly impacting the field of soft matter and are here to stay: (i) Supervised machine learning techniques – primarily kernel methods and neural network – will likely contribute to more accurate and transferable force fields; (ii) high-throughput methods have already pushed the boundaries of conformational sampling and are likely to affect the systematic screening of compounds and materials; and (iii) Bayesian inference provides a conceptually appealing framework to combine

simulation data and physical laws and symmetries. Another notable method of interest is unsupervised machine learning, which looks for features/structure in "unlabeled" datasets, such as clustering or dimensionality reduction techniques (Fisher et al. 2014). The rapid ongoing developments of unsupervised machine learning are likely to significantly affect computer simulations in the years to come.

# References

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104(13):136403

Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145(17):170901

Bereau T (2018) Example: ML model of Hirshfeld ratios. https://gitlab.mpcdf.mpg.de/trisb/handbook_example. Accessed 28 Feb 2018

Bereau T, Kremer K (2015) Automated parametrization of the coarse-grained martini force field for small organic molecules. J Chem Theory Comput 11(6):2783–2791

Bereau T, Andrienko D, von Lilienfeld OA (2015) Transferable atomic multipole machine learning models for small organic molecules. J Chem Theory Comput 11(7):3225–3233

Bereau T, Andrienko D, Kremer K (2016) Research update: computational materials discovery in soft matter. APL Mater 4(5):053101

Bereau T, DiStasio RA Jr, Tkatchenko A, von Lilienfeld OA (2018) Non-covalent interactions across organic and biological subsets of chemical space: physics-based potentials parametrized from machine learning. J Chem Phys 147(24):241706

Bowman GR, Pande VS, Noé F (Eds) (2013) An introduction to Markov state models and their application to long timescale molecular simulation, Advances in Experimental Medicine and Biology 797. Springer, Dordrecht (NL)

Chiavazzo E, Covino R, Coifman RR, Gear CW, Georgiou AS, Hummer G, Kevrekidis IG (2017) Intrinsic map dynamics exploration for uncharted effective free-energy landscapes. Proc Natl Acad Sci 114(28):E5494–E5503

Chmiela S, Tkatchenko A, Sauceda HE, Poltavsky I, Schütt KT, Müller KR (2017) Machine learning of accurate energy-conserving molecular force fields. Sci Adv 3(5):e1603015

Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12(3):191–201

Deringer VL, Csányi G (2017) Machine learning based interatomic potential for amorphous carbon. Phys Rev B 95(9):094203

Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, Vinyals O, Kearnes S, Riley PF, von Lilienfeld OA (2017) Machine learning prediction errors better than DFT accuracy. arXiv e-prints arXiv:170205532

Ferguson AL (2017) Bayeswham: a Bayesian approach for free energy estimation, reweighting, and uncertainty quantification in the weighted histogram analysis method. J Comput Chem 38(18):1583–1605

Ferguson AL, Panagiotopoulos AZ, Debenedetti PG, Kevrekidis IG (2011) Integrating diffusion maps with umbrella sampling: application to alanine dipeptide. J Chem Phys 134(13):04B606

Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. Phys Rev Lett 63(12):1195

Fisher DH, Pazzani MJ, Langley P (eds) (2014) Concept formation: knowledge and experience in unsupervised learning. Morgan Kaufmann Series in Machine Learning, San Mateo (CA)

Glielmo A, Sollich P, De Vita A (2017) Accurate interatomic force fields via machine learning with covariant kernels. Phys Rev B 95(21):214302

Halgren TA (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. J Am Chem Soc 114(20):7827–7843

Halgren TA, Damm W (2001) Polarizable force fields. Curr Opin Struct Biol 11(2):236–242

Huan TD, Batra R, Chapman J, Krishnan S, Chen L, Ramprasad R (2017) A universal strategy for the creation of machine learning-based atomistic force fields. npj Comput Mater 3(1):37

Huang B, von Lilienfeld O (2016) Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity. J Chem Phys 145(16):161102–161102

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. Apl Mater 1(1):011002

John S (2016) Many-body coarse-grained interactions using gaussian approximation potentials. arXiv preprint arXiv:161109123

Kukharenko O, Sawade K, Steuer J, Peter C (2016) Using dimensionality reduction to systematically expand conformational sampling of intrinsically disordered peptides. J Chem Theory Comput 12(10):4726–4734

Li Y, Li H, Pickard FC IV, Narayanan B, Sen FG, Chan MK, Sankaranarayanan SK, Brooks BR, Roux B (2017) Machine learning force field parameters from ab initio data. J Chem Theory Comput 13(9):4492–4503

Li Z, Kermode JR, De Vita A (2015) Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. Phys Rev Lett 114(9):096405

Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. J Comput Chem 25(13):1584–1604

Maple JR, Dinur U, Hagler AT (1988) Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. Proc Natl Acad Sci 85(15):5350–5354

Marrink SJ, Tieleman DP (2013) Perspective on the MARTINI model. Chem Soc Rev 42(16):6801–6822

Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH (2007) The martini force field: coarse grained model for biomolecular simulations. J Phys Chem B 111(27):7812–7824

Menichetti R, Kanekal KH, Kremer K, Bereau T (2017a) In silico screening of drug-membrane thermodynamics reveals linear relations between bulk partitioning and the potential of mean force. J Chem Phys 147(12):125101

Menichetti R, Kremer K, Bereau T (2017b) Efficient potential of mean force calculation from multiscale simulations: solute insertion in a lipid membrane. Biochem Biophys Res Commun. 498:282–287. https://doi.org/10.1016/j.bbrc.2017.08.095

Morawietz T, Singraber A, Dellago C, Behler J (2016) How Van der Waals interactions determine the unique properties of water. Proc Natl Acad Sci 113:8368–8373

Neale C, Bennett WD, Tieleman DP, Pomès R (2011) Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers. J Chem Theory Comput 7(12):4175–4188

Noé F (2008) Probability distributions of molecular observables computed from Markov models. J Chem Phys 128(24):244103

Noid W (2013) Perspective: coarse-grained models for biomolecular systems. J Chem Phys 139(9):09B201_1

Olsson S, Wu H, Paul F, Clementi C, Noé F (2017) Combining experimental and simulation data of molecular processes via augmented Markov models. Proc Natl Acad Sci 114(31): 8265–8270

Perez A, MacCallum JL, Dill KA (2015) Accelerating molecular simulations of proteins using Bayesian inference on weak information. Proc Natl Acad Sci 112(38):11846–11851

Perez A, Morrone JA, Dill KA (2017) Accelerating physical simulations of proteins by leveraging external knowledge. Wiley Interdiscip Rev Comput Mol Sci 7:e1309

Peter C, Kremer K (2010) Multiscale simulation of soft matter systems. Faraday Discuss 144:9–24

Plattner N, Doerr S, De Fabritiis G, Noe F (2017) Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. Nat Chem 9:1005–1011

Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66:27–85

Rabiner L, Juang B (1986) An introduction to hidden Markov models. IEEE ASSP Mag 3(1):4–16

Ramakrishnan R, von Lilienfeld OA (2017) Machine learning, quantum chemistry, and chemical space. Rev Comput Chem 30:225–256

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning, vol 1. MIT Press, Cambridge (MA)

Rudzinski JF, Kremer K, Bereau T (2016) Communication: consistent interpretation of molecular simulation kinetics using Markov state models biased with external information. J Chem Phys 144(5):051102

Rühle V, Junghans C, Lukyanov A, Kremer K, Andrienko D (2009) Versatile object-oriented toolkit for coarse-graining applications. J Chem Theory Comput 5(12):3211–3223

Rupp M, Tkatchenko A, Müller KR, Von Lilienfeld OA (2012) Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 108(5):058301

Schiilkopf B (2001) The kernel trick for distances. In: *Advances in neural information processing systems. Proceedings of the 2000 conference*, vol 13. MIT Press, Cambridge (MA), p 301

Shaw DE, Grossman J, Bank JA, Batson B, Butts JA, Chao JC, Deneroff MM, Dror RO, Even A, Fenton CH et al (2014) Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In: Proceedings of the international conference for high performance computing, networking, storage and analysis. IEEE Press, New Orleans, pp 41–53

Sodt AJ, Sandar ML, Gawrisch K, Pastor RW, Lyman E (2014) The molecular structure of the liquid ordered phase of lipid bilayers. J Am Chem Soc 136(2):725

Stroet M, Koziara KB, Malde AK, Mark AE (2017) Optimization of empirical force fields by parameter space mapping: a single-step perturbation approach. J Chem Theory Comput 13:6201–6212

Tetko IV, Tanchuk VY, Villa AEP (2001) Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. J Chem Inf Comput Sci 41(5):1407–1421

Tkatchenko A, DiStasio RA Jr, Car R, Scheffler M (2012) Accurate and efficient method for many-body van der Waals interactions. Phys Rev Lett 108(23):236402

Van Vleet MJ, Misquitta AJ, Stone AJ, Schmidt JR (2016) Beyond Born–Mayer: improved models for short-range repulsion in ab initio force fields. J Chem Theory Comput 12(8):3851–3870

Voth GA (2008) Coarse-graining of condensed phase and biomolecular systems. CRC Press, Boca Raton

Wang W, Donini O, Reyes CM, Kollman PA (2001) Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct 30(1):211–243

# Hydrodynamics in Motile Active Matter

# 68

Roland G. Winkler and Gerhard Gompper

## Contents

### Abstract

Hydrodynamic interactions determine the individual and collective behavior of nano- to micrometer size active objects such as swimming bacteria, sperm, algae, and synthetic colloidal microswimmers. Based on the Navier-Stokes equations of hydrodynamics, the major contributions to the flow field of a swimmer in a

R. G. Winkler (✉) · G. Gompper
Institute for Advanced Simulation and Institute for Complex Systems, Forschungszentrum Jülich, Jülich, Germany
e-mail: r.winkler@fz-juelich.de; g.gompper@fz-juelich.de

Newtonian fluid are presented. The propulsion of beating and rotating filaments is shown to emerge as consequence of the distinct friction coefficients for parallel and perpendicular motion of the filament. Hydrodynamic interactions with a wall lead to a preferred alignment of a swimmer adjacent to a wall. Moreover, the rotational motion of a flagellar bundle of swimming bacteria combined with the counterrotation of the cell body leads to circular trajectories on a surface, where the handedness depends on the wall slip. Even more, the collective behavior of active matter is determined by hydrodynamic interactions, which is illustrated by cilia synchronization and the squirmer model for microswimmers.

# 1    Introduction

Active matter, whose agents consume internal energy or extract energy from the environment to propel themselves through a fluid, and are thus far from thermal equilibrium, is omnipresent in nature. Examples on the microscale range from an uncountable number of bacteria in soil or living in symbiosis with humans, spermatozoa in their attempt to fertilize an ovum, or algae harvesting sunlight in ponds and the ocean. Nowadays, synthetic active systems have been designed, which are powered by phoretic processes, e.g., thermophoresis or diffusiophoresis (Bechinger et al. 2016). In any case, the microswimmer is embedded in a fluid, and the fluid plays a decisive role for the propulsion itself as well as the collective behavior (Lauga and Powers 2009; Yeomans et al. 2014; Elgeti et al. 2015; Zöttl and Stark 2016; Winkler 2016). The physics ruling swimming on the micrometer scale is very different from that applying to swimming in the macro-world, although certain propulsion strategies are reminiscent of those on a macro-scale – bacteria, such as *Escherichia coli*, are propelled by rotating flagella, sperm perform a snakelike motion, and algae, such as *Chlamydomonas reinhardtii*, apply a breaststroke-type beating pattern. However, swimming at the micrometer scale is swimming at low-Reynolds numbers (Purcell 1977), where viscous damping by far dominates over inertia. Hence, swimming concepts of the high-Reynolds number macro-world are ineffective on small scales. In the evolutionary process, microorganisms acquired propulsion strategies, which successfully overcome and even exploit viscous drag.

Bacteria, sperm, or algae use flagella – filamentous structures protruding from their bodies – for their propulsion (cf. Fig. 1 for an illustration). In fact, eukaryotic flagella are very different from prokaryotic ones, which is manifested in the differing propulsion strategies (Elgeti et al. 2015). However, in any case, the thrust force emerges by the difference in the hydrodynamic friction of a (long) slender body parallel and perpendicular to the body major axis. Thereby, the flow field far from the swimmer is usually dominated by a "force dipole" and decays similarly with distance, independent of the propulsion mechanism, since a microswimmer is force- and torque-free. At walls, surface hydrodynamic interactions lead to a propulsion-dependent preferred alignment of a microswimmer (Spagnolie and Lauga 2012) or circular trajectories (Lauga et al. 2006; Di Leonardo et al. 2011; Hu et al. 2015a). Under shear flow, the hydrodynamic force-dipole flow field substantially affects
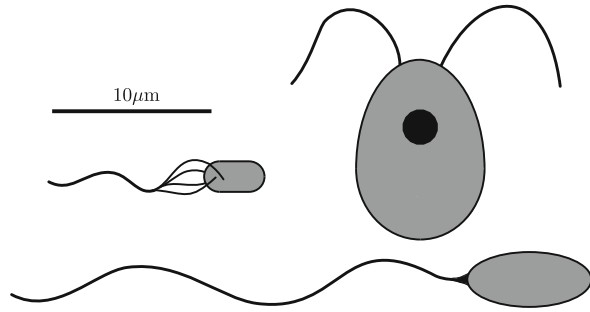
the overall viscosity (Saintillan 2010), and for suspensions of pushers, e.g., *E. coli* bacteria, "superfluidlike" behavior has been observed, where the viscous resistance to shear vanishes (López et al. 2015).

Active matter exhibits fascinating emergent collective phenomena. In nature, microswimmers can reach astonishing densities. Sperm cells are released by the millions to compete in the run for the egg, and biofilms are made up of billions of bacteria. Coordinated motion is exploited by spermatozoa of some species by self-assembling into unique train-like aggregates of hundreds or thousands of cells and thereby significantly increased sperm motility in a viscous environment (Sivinski 1984; Moore and Taggart 1995). Flagellated bacteria exhibit a particular mode of motion, where they migrate collectively over surfaces and are able to form stable aggregates, which can become highly motile (Heinrichsen 1978; Copeland and Weibel 2009; Kearns 2010). Here, cooperativity reaches a new level, and bacteria exhibit highly organized movements with remarkable large-scale patterns such as networks, complex vortices, or swarms (Copeland and Weibel 2009; Wensink et al. 2012). These type of patterns are remarkably similar to patterns appearing for other active matter systems such as schools of fish, flocks of birds, mammalian herds, or crowds of humans (Vicsek and Zafeiris 2012; Elgeti et al. 2015; Popkin 2016).

The various aspects touched above illustrate the fundamental importance of hydrodynamics for microscopic active matter ranging from swimming of individuals to large-scale collective migration. In the following, the low-Reynolds number aspects relevant for microswimmers will be briefly summarized, and various hydrodynamic phenomena will be presented.

## 2   Low-Reynolds Number Hydrodynamics

### 2.1   Equations of Motion

Typically, the dynamics of the (isothermal) incompressible fluid flow field surrounding a microswimmer is described by the Navier-Stokes equations

$$\rho \left( \frac{\partial}{\partial t} \boldsymbol{v} + (\boldsymbol{v} \cdot \nabla) \, \boldsymbol{v} \right) = -\nabla p + \eta \nabla^2 \boldsymbol{v} + \boldsymbol{f} \,, \qquad \nabla \cdot \boldsymbol{v} = 0, \qquad (1)$$

where $\boldsymbol{v}(\boldsymbol{r}, t)$, $p(\boldsymbol{r}, t)$, and $\boldsymbol{f}(\boldsymbol{r}, t)$ are the velocity, pressure, and volume-force density fields, respectively. At small Reynolds numbers $Re = \rho u L / \eta \ll 1$, where $\rho$ is the fluid mass density, $u$ the characteristic velocity, $L$ the size of the microswimmer, and $\eta$ the fluid viscosity, the inertia terms on the left-hand side of Eq. (1) can be neglected, and the equations reduce to the Stokes or creeping flow equations

$$\nabla p(\boldsymbol{r}) - \eta \nabla^2 \boldsymbol{v}(\boldsymbol{r}) = \boldsymbol{f}(\boldsymbol{r}) \,, \qquad \nabla \cdot \boldsymbol{v} = 0. \qquad (2)$$

For illustration, the Reynolds number in water of a swimmer of length $L = 10 \, \mu\text{m}$, a velocity of $u = 50 \mu\text{m/s}$, and the kinematic viscosity $\nu = \eta / \rho = 10^{-6} \text{m}^2/\text{s}$ is $Re \approx 10^{-3}$. The Stokes equation (2) is linear and time independent. The consequences of this intrinsic symmetry under time reversal for microswimmers undergoing periodic shape changes were first expressed by Purcell (1977) and are now known as "scallop theorem", which can be stated as: if the shape changes displayed by a swimmer are identical when viewed in reverse order (time reversal symmetry), it will generate an oscillatory, but no directed motion (Purcell 1977; Lauga and Powers 2009; Yeomans et al. 2014; Elgeti et al. 2015). Thus, just by opening and closing its two shells, a mussel (scallop) cannot move forward at $Re \ll 1$. Microswimmers developed various strategies to beat the scallop theorem. Aside from many (elastic) degrees of freedom, they use specific propulsion mechanisms which are not time reversible – bacteria such as *E. coli* are propelled by rotating helical flagella bundles, sperm use sinusoidal bending waves propagating from head to tail, and algae, e.g., *Chlamydomonas*, use a nonreciprocal stroke pattern.

## 2.2 Solution of Stokes Equation

The linear Stokes equations (2) are easily solved analytically for an unbounded fluid. The respective fluid velocity field is

$$\boldsymbol{v}(\boldsymbol{r}) = \int \mathbf{Q}(\boldsymbol{r} - \boldsymbol{r}') \boldsymbol{f}(\boldsymbol{r}') \, d^3 r' \,, \qquad Q_{\alpha\alpha'}(\boldsymbol{r}) = \frac{1}{8\pi \eta r} \left[ \delta_{\alpha\alpha'} + \frac{r_\alpha r_{\alpha'}}{r^2} \right], \qquad (3)$$

where $\mathbf{Q}(\boldsymbol{r})$ is the well-known Oseen tensor, with the Cartesian components $Q_{\alpha\alpha'}$ ($\alpha, \alpha' \in \{x, y, z\}$) and $r = |\boldsymbol{r}|$ (Kim and Karrila 1991; Dhont 1996). The Oseen tensor, also denoted as Stokeslet, shows that hydrodynamic interactions are long ranged, with a $1/r$ decay like the Coulomb potential, and are anisotropic due to the incompressibility of the fluid. The Oseen tensor is the Green's function of the Stokes equation (2), which is evident, when the point force $\boldsymbol{f}(\boldsymbol{r}) = f_0 \delta(\boldsymbol{r}) \boldsymbol{e}$ in the direction $\boldsymbol{e}$ ($|\boldsymbol{e}| = 1$) is inserted. Then, Eq. (3) yields

$$v(r) = \frac{f_0}{8\pi\eta r}\left[e + \frac{(r\cdot e)r}{r^2}\right]. \tag{4}$$

The magnitude of the flow field is twice larger in the force direction than perpendicular to it.

## 2.3  Microswimmer Flow Field

Most microswimmers move autonomously, with no external force or torque applied, and hence the total force/torque of the swimmer on the fluid and *vice versa* vanishes. In the simplest case, which actually applies to many microswimmers like bacteria, spermatozoa, or algae, the far-field hydrodynamics (at distances from the swimmer much larger than its size) can well be described by a force dipole (Lauga and Powers 2009; Ishikawa 2009). This has been confirmed experimentally for *E. coli* (Drescher et al. 2010, 2011) and in simulations (Hu et al. 2015b). The flow field of *Chlamydomonas* is well reproduced by three Stokeslets (Drescher et al. 2010).

Mathematically, the flow field $v(r-r_0)$ of a hydrodynamic force dipole located at $r_0$ follows by a superposition of two Stokeslets (4) with opposite forces $f_0 = \pm f_0 e$ of equal magnitude at $r_0 \pm l/2$, where $l = le$ and $l$ is the distance between the Stokeslets. Taylor expansion to leading order in $|l|/|r - r_0|$ yields

$$v(r) = \frac{P}{8\pi\eta}v^{FD}(r)\,,\quad v^{FD}(r) = \frac{r}{r^3}\left[-1 + 3\frac{(r\cdot e)^2}{r^2}\right], \tag{5}$$

where $P = \pm f_0 l$ is the dipole strength. Note that the flow field of a force dipole decays as $1/r^2$ from the center of the dipole, faster than the force monopole or Stokeslet Eq. (3). The flow fields of hydrodynamic dipoles are shown in Fig. 2. In two dimensions, there are two inflow (left, right) and two outflow (top, bottom) regions, which are separated by the separatrices $z = \pm\sqrt{2}x$. In three dimensions, the outflow region is a cone.

Two classes of dipole swimmers can be distinguished. A swimmer with its "motor" in the back, and a passive body dragging along the surrounding fluid in front, creates a "pusher" flow field (cf. Fig. 2 (left)). Similarly, a swimmer with its "motor" in front, and the passive body dragging along the fluid behind, develops a "puller" flow field. This field follows by inversion of the arrows in Fig. 2 (left), i.e., the flow fields of pushers and pullers look similar but with opposite flow directions. This has important consequences for the interactions between swimmers and of swimmers with walls, as will be explained below.

The dipolar flow field and higher-order multipoles follow by a systematic expansion of the Oseen tensor in Eq. (3) (Kim and Karrila 1991; Pozrikidis 1992; Spagnolie and Lauga 2012). For a sphere of radius $R$, the swimmer far field up to order $\mathcal{O}(r^{-4})$ is dominated by the force dipole (FD) (Eq. (5)), source dipole (SD), force quadrupole (FQ), source quadrupole (SQ), and rotlet dipol (RD) contributions
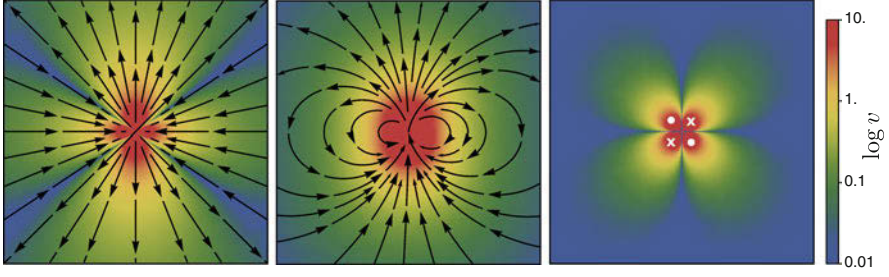
**Fig. 2** (Left) Flow lines in the far-field of a hydrodynamic force dipole (5) and (Middle) a source dipole (7) oriented along the vertical direction. (Right) The flow field of the rotlet dipole (10) is rotational symmetric around the horizontal axis. The white crosses and bullets indicate the intersections of the flow lines with the plane

$$
\boldsymbol{v}(\boldsymbol{r}) = \kappa^{FD}\boldsymbol{v}^{FD}(\boldsymbol{r}) + \kappa^{SD}\boldsymbol{v}^{SD}(\boldsymbol{r}) + \kappa^{FQ}\boldsymbol{v}^{FQ}(\boldsymbol{r}) + \kappa^{SQ}\boldsymbol{v}^{SQ}(\boldsymbol{r})
$$
$$
+ \kappa^{RD}\boldsymbol{v}^{RD}(\boldsymbol{r}) + \mathcal{O}(r^{-4}), \tag{6}
$$

where

$$
\boldsymbol{v}^{SD}(\boldsymbol{r}) = -\frac{1}{r^3}\left(-\boldsymbol{e}_z + \frac{3z\boldsymbol{r}}{r^2}\right), \tag{7}
$$

$$
\boldsymbol{v}^{FQ}(\boldsymbol{r}) = \frac{1}{r^3}\left[\left(1 - \frac{3z^2}{r^2}\right)\boldsymbol{e}_z + \left(\frac{15z^3}{r^4} - 9\frac{z}{r^2}\right)\boldsymbol{r}\right], \tag{8}
$$

$$
\boldsymbol{v}^{SQ}(\boldsymbol{r}) = \frac{3}{r^4}\left(\frac{5z^2\boldsymbol{r}}{r^3} - \frac{2z\boldsymbol{e}_z + \boldsymbol{r}}{r}\right), \tag{9}
$$

$$
\boldsymbol{v}^{RD}(\boldsymbol{r}) = \frac{3z\boldsymbol{e}_z \times \boldsymbol{r}}{r^5}, \tag{10}
$$

which decay like $r^{-2}, r^{-3}, r^{-4}$, and $r^{-3}$, respectively (Spagnolie and Lauga 2012). Note that in Eq. (7) the swimming direction $\boldsymbol{e}$ points along the positive $z$-axis, i.e., $\boldsymbol{e} \equiv \boldsymbol{e}_z$. The various factors $\kappa$ account for the strength of the respective multipole, where $\kappa^{FD} = P/8\pi\eta$, $\kappa^{SD} = -v_0 R^3/2$, and $\kappa^{SQ} = 3PR^3/8\pi\eta$.

Aside from the force-dipole term, most relevant for microswimmer are the source-dipole term $(1/r^3)$ due to the volume of the swimmer and the rotlet-dipole term $(1/r^3)$, e.g., for *E. coli* bacteria by the opposite rotation of the cell body and the flagella bundle.

## 3 Swimming Due to Flagellar Motion

### 3.1 Friction of Slender Body

Many microorganisms are propelled in a fluid by beating or rotating a flagellum. Thereby, a swimmer exploits the viscous frictional properties of the fluid environment, specifically, the anisotropic friction of a slender body. The frictional anisotropy can be demonstrated for a long and thin rod of radius $R$ and length $L$. Considering the rod as composed of a sequence of beads with no-slip boundary conditions, the beat velocity is equal to the fluid velocity of Eq. (3). Under the influence of an external constant force $\boldsymbol{F} = F\boldsymbol{e}$, the average velocity of a rod aligned along the $z$-axis of the reference system is (Elgeti et al. 2015)

$$\boldsymbol{v}_{\text{rod}} = F\frac{\boldsymbol{e} + (\boldsymbol{e}_z \cdot \boldsymbol{e})\boldsymbol{e}_z}{4\pi\eta L^2} \int_{2R}^{L} \frac{L-s}{s} ds \qquad (11)$$

in the limit of a continuous rod. The lower cutoff of the integral excludes a region of the thickness of the rod. Because $(\boldsymbol{e}_z \cdot \boldsymbol{e})\boldsymbol{e}_z$ is 1 and 0 for parallel and perpendicular orientation of the force with the rod axis, respectively, evaluation of the integral and the relation $\boldsymbol{F} = \zeta_\parallel \boldsymbol{v}_\parallel + \zeta_\perp \boldsymbol{v}_\perp$ yields

$$\zeta_\perp = 2\zeta_\parallel, \qquad \zeta_\perp = \frac{4\pi\eta L}{\ln(L/2R)} \qquad (12)$$

in the asymptotic limit of a long rod. Hence, pulling a rod along its axis is easier than perpendicular to it. The logarithmic divergence is a result of the long-range nature of hydrodynamic interactions between different parts of the rod. Thus, hydrodynamic interactions reduce the friction coefficient compared to that of a rod of hydrodynamically noninteracting beads, where $\zeta_\perp = \zeta_\parallel \sim L$. Corrections of the friction coefficients for a more precise account of hydrodynamics for a cylinder have been calculated (Tirado et al. 1984).

### 3.2 Propulsion by Beating Flagella

The time-dependent shape of a sinusoidally beating flagellum with a planer beat ($xz$-plane) $z(x, t)$ and its local velocity $v_z(x, t)$ at the position $x$ along its contour are described by

$$z(x, t) = A\sin(kx - wt), \quad v_z(x, t) = \frac{\partial z}{\partial t} = -A\omega\cos(kx - \omega t), \qquad (13)$$

where $A$ is the amplitude, $\omega$ the frequency, and $k$ the wave number. Decomposing the velocity $\boldsymbol{v}(x, t) = (0, 0, v_z(x, t))^T$ into a component parallel $\boldsymbol{v}_\parallel = (\boldsymbol{v} \cdot \boldsymbol{t})\boldsymbol{t}$ and perpendicular $\boldsymbol{v}_\perp = \boldsymbol{v} - \boldsymbol{v}_\parallel$ to the local tangent vector $\boldsymbol{t} \sim (1, 0, Ak\cos(kx - \omega t))^T$ yields

$$\boldsymbol{v}_\parallel = -\frac{A^2\omega k \cos^2(kx - \omega t)}{1 + A^2 k^2 \cos^2(kx - \omega t)}\, \boldsymbol{t}. \tag{14}$$

The separation $\boldsymbol{F} = \zeta_\parallel \boldsymbol{v}_\parallel + \zeta_\perp \boldsymbol{v}_\perp$ gives the average force of the flagellum in the swimming direction

$$F_z = (\zeta_\parallel - \zeta_\perp)\frac{1}{L}\int \frac{A^2\omega k \cos^2(kx - \omega t)}{1 + A^2 k^2 \cos^2(kx - \omega t)}\, dx, \tag{15}$$

while the average force in the perpendicular direction vanishes. For small beating amplitudes, Eq. (15) can easily be integrated, which yields the average propulsion force

$$F_z = \frac{1}{2}(\zeta_\parallel - \zeta_\perp)A^2\omega k, \tag{16}$$

and the swimming velocity, $v_{\text{flag}} \approx F_z/\zeta_\parallel$, (Gray and Hancock 1955)

$$v_{\text{flag}} = -\frac{1}{2}\left(\frac{\zeta_\perp}{\zeta_\parallel} - 1\right)A^2\omega k\,. \tag{17}$$

This simplified calculation shows several important aspects of flagellar propulsion. First, swimming is only possible due to the frictional anisotropy, i.e., $\zeta_\parallel \neq \zeta_\perp$. Second, for a traveling wave in the positive $x$-direction, the flagellum moves in the negative $x$-direction, i.e., movement is opposite to the direction of the traveling wave. Third, the swimming velocity increases linearly with the beating frequency $\omega$ and the wave vector $k$ but quadratically with the beating amplitude $A$. And finally, the swimming velocity is independent of the fluid viscosity for a given beating amplitude.

### 3.3 Propulsion by Helical Flagella

Propulsion by rotation of helical flagella can also be illustrated by resistive force theory. Rotation of a rodlike segment in the direction $\boldsymbol{v}'$, where $v' = |\boldsymbol{v}'| = R_h\Omega$ (cf. Fig. 3), with $R_h$ the helix radius and $\Omega$ its rotation frequency, yields the thrust force contribution $F_T$ and the torque $M_z$ (Lauga and Powers 2009)

$$F_T = (\zeta_\parallel - \zeta_\perp)v' \cos\vartheta \sin\vartheta \approx (\zeta_\parallel - \zeta_\perp)\vartheta\, R_h\Omega, \quad M_z = \zeta_\perp R_h v' = \zeta_\perp R_h^2\Omega, \tag{18}$$

where $\vartheta \ll 1$ is assumed. Hence, the relation between force, torque, and translational and rotational velocity is
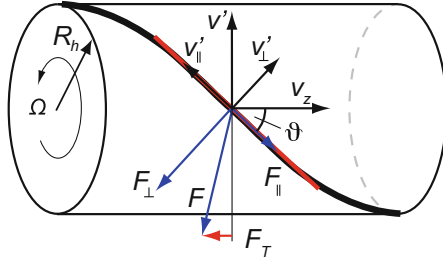
**Fig. 3** Helical segment moving in a viscous fluid. Only half of a helical pitch is shown. The drag-based thrust force $F_T$ appears by the rotation of the red rodlike segment in the direction $v'$. The orientation angle $\vartheta$ is related with the pitch angle by $\pi/2 - \vartheta$. (From Elgeti et al. 2015)

$$\begin{pmatrix} F_z \\ M_z \end{pmatrix} = \begin{pmatrix} \zeta_\parallel & -(\zeta_\perp - \zeta_\parallel)\vartheta R_h \\ -(\zeta_\perp - \zeta_\parallel)\vartheta R_h & \zeta_\perp R_h^2 \end{pmatrix} \begin{pmatrix} v_z \\ \Omega \end{pmatrix}, \tag{19}$$

with $F_z = \zeta_\parallel v_z$ in case of no helix rotation. For a spherical cell body of radius $R_b$ and with the assumption $R_b \ll L$, the frictional body force $F_b$ and the body torque $M_b$ are

$$F_b = \zeta_b v_z, \quad M_b = -\zeta_r^b \omega_b, \tag{20}$$

where $\zeta_b = 6\pi \eta R_b$ and $\zeta_r^b = 8\pi \eta R_b^3$ are the translation and rotational friction coefficients. The helix is driven by a rotary motor with the frequency $\Omega_m$ relative to the body. In response, the helix and body rotate with the frequencies $\Omega$ and $\Omega_b$. These frequencies are related by $\Omega + \Omega_b = \Omega_m$. Since the whole bacterium is force- and torque-free, i.e., $F_z + F_b = 0$ and $M_z + M_b = 0$, its swimming velocity is obtained as

$$v_z \approx \vartheta \left( \frac{\zeta_\perp}{\zeta_\parallel} - 1 \right) \frac{\zeta_r^b}{\zeta_\perp R_h} \Omega_m. \tag{21}$$

The friction coefficient $\zeta_b$ does not appear, since $\zeta_\parallel \gg \zeta_b$ ($L \gg R_b$) is assumed. Evidently, swimming is again – as in the sperm case – only possible due to frictional anisotropy. Moreover, $v_z$ depends linearly on the body rotational friction coefficient. Hence, without body, the bacterium could not swim. Due to the approximation $\vartheta \ll 1$, $v_z$ depends linearly on the orientation angle $\vartheta$. Changing the handedness of the helix leads to a change of the swimming direction.

Note that a helix driven by an external torque also moves forward; however, it is not torque-free, and therefore is not an autonomous swimmer.

## 3.4    Bacteria Swimming

A wide variety of bacteria exploits the propulsion strategy described in Sect. 3.3. Different species possess various numbers and differing arrangements of flagella. According to the arrangement, flagellated bacteria are classified as monotrichous bacteria with a single flagellum only, lophotrichous bacteria with multiple flagella located at a particular spot on their surface, amphitrichous bacteria with a single flagellum on each of the two opposite ends, and peritrichous bacteria which are covered by multiple flagella pointing in all directions. Prominent examples of peritrichous bacteria are *E. coli*, *Salmonella typhimurium*, *Rhizobium lupini*, or *Proteus mirabilis*, to name just a few. A flagellum is rotated by a motor complex, which consists of several proteins, and is anchored in the bacterial cell wall (Berg 2003). Bacteria like *E. coli* swim in a "run-and- tumble" motion (Berg 2003). In the "run" phase, the helical flagella are left-handed, and they rotate counterclockwise. The flagella form a bundle, and the bacterium moves forward in a direction determined by its long axis. At the beginning of the "tumble" phase, a flagellum rotational direction is reverted to clockwise. The flagellum leaves the bundle, which implies a random reorientation of the bacterium. The reversal of the rotational direction is accompanied by a change of the helical handedness from left-handed to right-handed, and the flagellum undergoes a polymorphic transition, i.e., assumes a different pitch and radius (Calladine 1975; Macnab 1977). At the end of the tumbling phase, all flagella start to rotate again in the same counterclockwise direction, the bundle reforms, and the bacterium returns to a directional motion. The flagella of bacteria like *Rhizobium meliloti* or *Rhizobium lupini* are only capable of limited polymorphic transitions, and their motors are unidirectional (Platzer et al. 1997). These bacteria modulate the rotation speed of individual motors to induce tumbling.

Since bacterial cells are force- and torque-free, the rotational motion of the flagellum bundle leads to a counterrotation of the cell body, i.e., swimming bacteria possess a rotlet dipole (cf. Eq. (10)). This has consequences for their hydrodynamic interactions, specifically with surfaces and interfaces.

The flow field of an *E. coli* bacterium obtained from experiment and simulations is presented in Fig. 4. In both cases, the far field is well described by the force-dipole field of Eq. (5) (Drescher et al. 2011; Hu et al. 2015b).

## 4    Surface Interaction

## 4.1    Dipole Swimmer Near a Wall: Swimming with an Image

The swimming behavior of microorganisms is typically altered by the presence of nearby obstacles or boundaries. In fact, most bacteria in nature live on surfaces, e.g., in biofilms (Copeland and Weibel 2009; Spagnolie and Lauga 2012). Correspondingly, attraction of such microorganisms to surfaces is of major importance
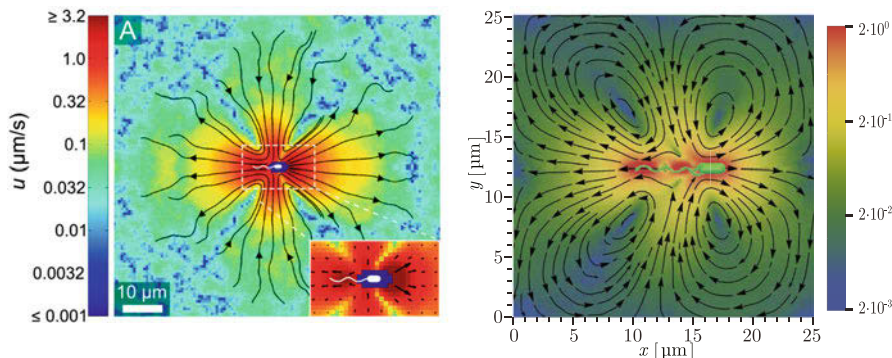
**Fig. 4** Flow field of *E. coli* bacteria from (left) experiment (Drescher et al. 2011) and (right) simulations (Hu et al. 2015b). In simulations, a system with periodic boundary conditions is considered, which yields closed flow lines in contrast to the flow lines of the experimental bulk system. The logarithmic color scheme (right) indicates the magnitude of the flow speed scaled by the bacterial swimming velocity. (From Drescher et al. 2011 and Hu et al. 2015b)

and determines their microbial activity. Part of the attraction originates from hydrodynamic interactions of the swimmer with the surface. This is easily illustrated by a force dipole in front of a (slip) surface. The flow field of such a dipole can be obtained by the image method known from electrostatics. Considering, for simplicity, a planar wall with slip boundary conditions (cf. Fig. 5), at $z = 0$, the velocity field $v_{w,z}^{FD}$ perpendicular to the surface vanishes identically, i.e., $v_{w,z}^{FD}(z = 0) \equiv 0$. The flow field is then given by

$$v_w^{FD}(r - r_0) = v^{FD}(r - \mathbf{r}_0; e) + v^{FD}(\mathbf{r} - r_0'; e'), \tag{22}$$

with $r_0 = (x_0, y_0, z_0)$, $r_0' = (x_0, y_0, -z_0)$, where $z_0 > 0$ and $e'$ the mirror image of $e$ with respect to the $z = 0$ plane. The dipole experiences a flow field and, hence, a force near the surface, which is determined by the hydrodynamic interactions between the dipole and its image. As a consequence, the dipole swimmer is moving toward the surface with the velocity

$$v_{w,z}^{FD}(z_0) = -\frac{P}{32\pi\eta z_0^2}\left[1 - 3(e \cdot e_z)^2\right], \tag{23}$$

because $(e' \cdot e_z)^2 = (e \cdot e_z)^2$. The result shows that the hydrodynamic force decays as a dipole flow field quadratic with the distance from the wall. The exact solution for a no-slip wall (Berke et al. 2008) yields the same functional dependence on the angle and the wall distance as Eq. (23), only the numerical prefactor in Eq. (23) is smaller by a factor 2/3.

The direction of the flow field depends on the dipole moment and its orientation. The hydrodynamic force is attractive to the wall for pusher ($P > 0$, sperm or
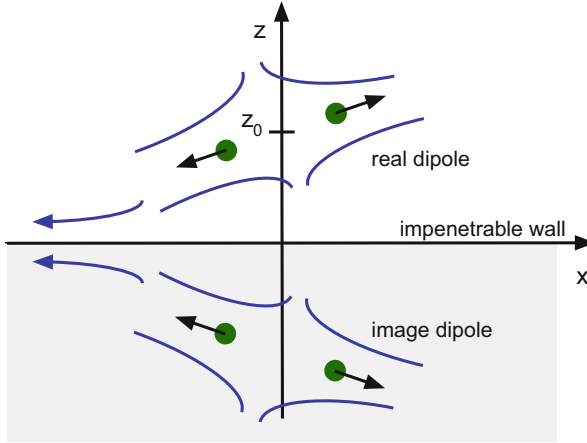
**Fig. 5** Schematic representation of a dipole swimmer (pusher) near a wall. An image dipole ensures the correct boundary conditions at the impenetrable slip wall

bacteria) as long as $\cos \vartheta = (e \cdot e_z) < 1/\sqrt{3}$. Hence, pushers aligned parallel to the surface experience an attractive force. However, for pullers ($P < 0$, *Chlamydomonas*) hydrodynamic interactions are repulsive when they swim parallel to the wall, but they are attractive, when their orientation is nearly perpendicular to the wall. Yet, the average of the wall-induced interaction over a population of randomly oriented microorganisms is exactly equal to zero in a 3D system, since $\int v_{w,z}^{FD}(z_0) \sin \vartheta \, d\vartheta = 0$. As a consequence, the surface-induced velocity (23) alone cannot explain surface accumulation for initially randomly oriented incoming swimming cells.

The surface-induced hydrodynamic flow field is inhomogeneous and, thus, exerts a torque on the cell, which leads to a preferred alignment. The corresponding rotation rate is given by (Berke et al. 2008):

$$\Omega_r(\vartheta, z) = -\frac{3P \cos \vartheta \sin \vartheta}{64\pi \eta z^3} \left( 1 + \frac{\gamma^2 - 1}{2(\gamma^2 + 1)} (1 + \cos^2 \vartheta) \right), \qquad (24)$$

where $\gamma$ is the aspect ratio of the anisotropic swimmer. Since $\gamma$ is typically larger than unity, the sign of $\Omega_r$ is determined by that of $P$ and the product $\cos \vartheta \sin \vartheta$. When $0 \leq \vartheta \leq \pi/2$ or $\pi \leq \vartheta \leq 3\pi/2$, for a pusher the product is positive, and the rotation is negative, leading to parallel alignment with the surface. For the other angles, the rotation is positive, which leads to alignment too. Consequently, all pushers are oriented parallel to a surface and are attracted by the flow field. Pullers ($P < 0$) align normal to the surface. As both pushers and pullers come closer to the surface, higher orders in the multipole expansion become important (Spagnolie and Lauga 2012).

## 4.2    Bacteria Swimming at Surfaces

As a consequence of the rotlet dipole of bacteria by counterrotation of cell body and flagella bundle, hydrodynamic interactions lead to circular trajectories of bacteria at surfaces (Lauga et al. 2006; Di Leonardo et al. 2011; Hu et al. 2015a). Thereby, clockwise and counterclockwise trajectories appear, governed by the respective no-slip or slip boundary condition (cf. Fig. 6) (Lauga et al. 2006; Di Leonardo et al. 2011; Hu et al. 2015a; Elgeti et al. 2015; Elgeti and Gompper 2016). Moreover, the slip length determines the curvature of the circle. In qualitative agreement with experiments and quantitative agreement with theory and simulations, the trajectory curvature can well be described by

$$\kappa = \kappa_\infty + \frac{\kappa_0 - \kappa_\infty}{1 + b/h} \qquad (25)$$

as function of the slip length $b$, where $\kappa_0 < 0$ and $\kappa_\infty > 0$ are the curvatures for the slip lengths $b = 0$ (no-slip) and $b = \infty$ (perfect slip), respectively, and $h$ is an effective gap size between of the cell body and the surface (Hu et al. 2015a).

As found experimentally, the radius of the circle depends on the size of the cell body and increases linearly with body size. This fact can be exploited to separate cells of different sizes. As suggested by Hu et al. (2015a), a patterned surface with alternating hydrophobic and hydrophilic stripes leads to a preferred diffusion parallel to the stripes for radii on the order of the stripe widths, whereas for larger radii isotropic diffusion is obtained.
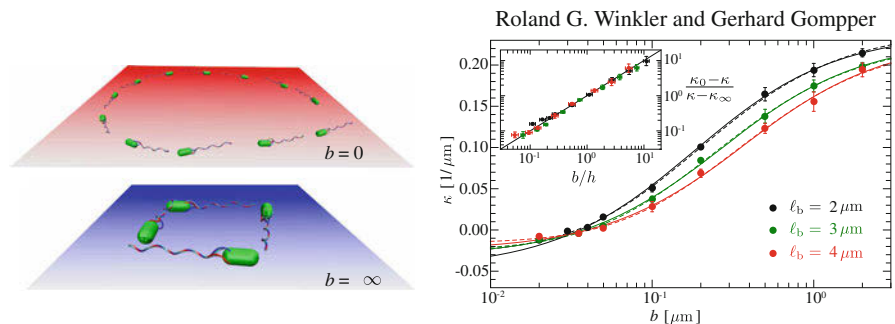


**Fig. 6** (Left) Counterclockwise and clockwise circular trajectories from hydrodynamic simulations of an *E. coli*-type bacterium swimming near homogeneous surfaces with different slip lengths $b$ as indicated. (Right) Effective curvatures of cells of various lengths, $l_b$, as function of the slip length. (From Hu et al. 2015a)

# 5 Squirmer: A Generic Model of Hydrodynamic Microswimmers

Generic models, which capture the essential swimming aspects, are crucial in theoretical studies of microswimmers. On the one hand, they help to unravel the relevant interaction mechanisms and, on the other hand, allow for the study of sufficiently large number of swimmers. A prominent example is the squirmer model (Lighthill 1952; Blake 1971). Originally, it was intended as a model for ciliated microswimmers, such as Paramecia. Nowadays, it is considered as a generic model for a broad class of microswimmers, ranging from diffusiophoretic particles to biological cells and has been applied to study collective effects in bulk, at surfaces, and in narrow slits (Ishikawa et al. 2006; Llopis and Pagonabarraga 2010; Zöttl and Stark 2014; Theers et al. 2016).

In its simplest form, a squirmer is represented as a spherical rigid colloid with a prescribed surface velocity. Restricting the surface velocity to be tangential, the slip velocity on the sphere surface can be expressed in terms of derivatives of Legendre polynomials, where the spherical squirmer is typically characterized by two modes only accounting for its swimming velocity ($B_1$) and its force dipole ($B_2$) (Ishikawa et al. 2006; Llopis and Pagonabarraga 2010). Explicitly, the leading contributions yield the slip velocity on the colloid surface (Ishikawa et al. 2006; Llopis and Pagonabarraga 2010; Theers et al. 2016)

$$\boldsymbol{v}_{sq} = (B_1 \sin \vartheta + B_2 \sin \vartheta \cos \vartheta)\boldsymbol{e}_\vartheta = B_1(\sin \vartheta + \beta \sin \vartheta \cos \vartheta)\boldsymbol{e}_\vartheta. \qquad (26)$$

The parameter $B_1 = 2v_0/3$ is related to the swimming velocity, $v_0$, and $\beta = B_2/B_1$ accounts for the force dipole. The angle $\vartheta$ is measured with respect to the propulsion direction in a body-fixed reference frame. Higher-order terms can easily be taken into account (Elgeti et al. 2015; Llopis and Pagonabarraga 2010). The term with $B_2$ (or $\beta$) distinguishes various propulsion patterns, namely, pushers ($\beta < 0$), pullers ($\beta > 0$), and neutral squirmers ($\beta = 0$), corresponding, e.g., to *E. coli*, *Chlamydomonas*, or *Volvox*, respectively.

The far field of a squirmer is well described by the flow fields of a force dipole (FD), a source dipole (SD), and a source quadrupole (SQ)

$$\boldsymbol{v}(\boldsymbol{r}) = \kappa^{FD}\boldsymbol{v}^{FD}(\boldsymbol{r}) + \kappa^{SD}\boldsymbol{v}^{SD}(\boldsymbol{r}) + \kappa^{SQ}\boldsymbol{v}^{SQ}(\boldsymbol{r}) + \mathcal{O}(r^{-5}), \qquad (27)$$

where the various terms are given in Eq. (5), (7), and (9).

The assumption of a spherical shape is adequate for swimmers like, e.g., *Volvox*; however, the shapes of other microswimmers (*E. coli, Chlamydomonas, Paramecium*) are nonspherical. Here, an extension of the squirmer concept to spheroidal objects has been proposed (Keller and Wu 1977; Theers et al. 2016). Figure 7 depicts flow fields of a spheroidal squirmer with the aspect ratio of two for the various kinds of dipolar terms in the laboratory and body-fixed reference frame. The near-field modifications by the finite-size swimmer is clearly visible in
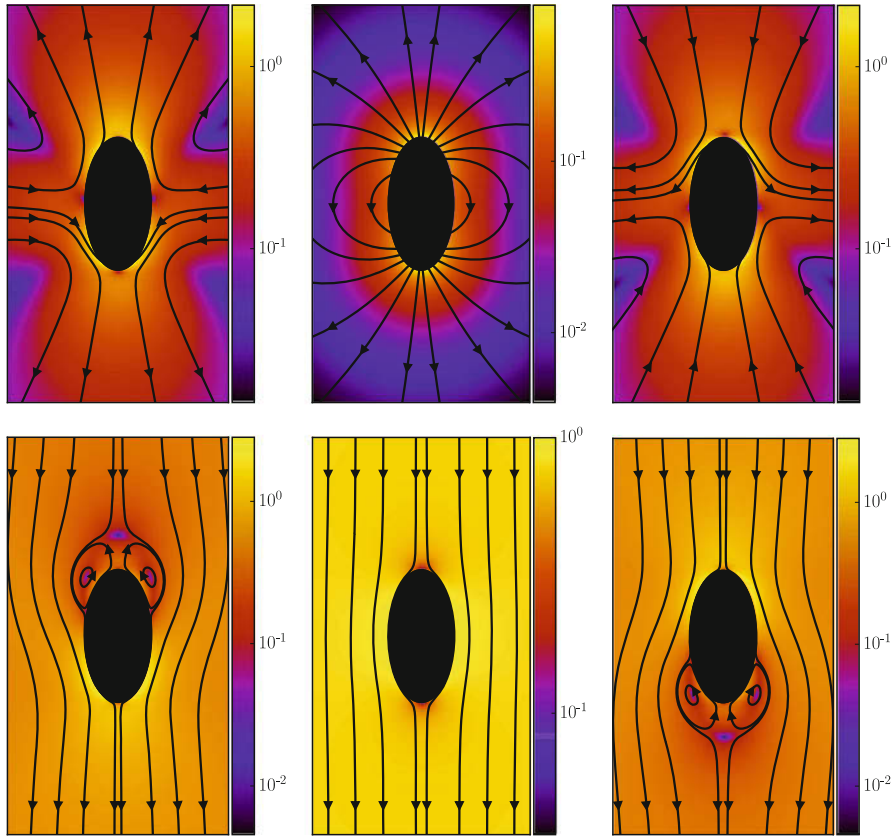
**Fig. 7** Flow streamlines of isolated spheroidal squirmers. The top row corresponds to the laboratory reference frame and the bottom row to the body-fixed reference frame. (Left) Flow field of a pusher ($\beta = -3$), (Middle) a neutral squirmer $\beta = 0$, and (Right) a puller ($\beta = 3$). The magnitude of the relative velocity $3v/2v_0$ is color coded logarithmically. (The puller figures are from Theers et al. 2016)

comparison with Fig. 2. Moreover, pusher and puller exhibit a stagnation point in front or back, respectively, in the body-fixed reference frame for $|\beta| > 1$.

## 6    Collective Phenomena

Collective phenomena governed by hydrodynamic interactions appear on the level of flagella or cilia as well as on the scale of the microswimmers themselves. Examples on the flagella scale are the synchronization of flagella rotation in the formation of bacteria bundles (Reichert and Stark 2005; Qian et al. 2009; Reigh et al. 2012) or the development of metachronal waves in arrays of beating cilia (Sleigh 1962; Elgeti and Gompper 2013).

The importance of hydrodynamic interactions on the collective dynamics of microswimmers is most easily demonstrated for squirmers. On the one hand, the effect of the force dipole (pusher, puller, neutral squirmer) is captured, and, on the other hand, both far- and near-field hydrodynamics is taken into account, and their relative importance can be elucidated. In accordance with bacteria in biofilms, the collective behavior of squirmers is either studied by strictly two-dimensional motion with three-dimensional hydrodynamic interactions or by swimmers confined in a narrow slit with a respective limitation of hydrodynamic interactions by the surfaces. In the latter case, the boundary interactions play a major role (Theers et al. 2018).

## 6.1    Cilia Synchronization: Metachronal Waves

Propulsion of unicellular and multicellular organisms by cilia is omnipresent. Thereby, fluid is pumped across their surface by anchored motile cilia (flagella) (Sleigh 1962). Moreover, in higher organisms and humans, cilia are involved in moving mucus in the lungs (Afzelius 1976), the emergence of the embryonic left-right asymmetry (Cartwright et al. 2004), and intercellular communication (Wang et al. 2006). Already in the 1960s, Sleigh (1962) observed that arrays of cilia beat neither randomly nor synchronously but in a wave pattern called a metachronal wave (MCW). Several theoretical models have been proposed to shed light onto the metachronal coordination by hydrodynamic interactions. A model of coupled rotating spheres placed near a no-slip wall proves useful in clarifying the diverse types of MCWs observed in nature (Brumley et al. 2012).

Deeper insight into coordinated beating is gained by simulations of anchored semiflexible filaments with a nonreciprocal beat, where a "trigger" mechanism switches between the power and recovery stroke. Simulations of an array of $60 \times 60$ cilia in a 3D explicit fluid yield metachronal waves emerging autonomously, despite the presence of significant noise (Elgeti and Gompper 2013). The beat pattern of an individual cilium can react to the surrounding fluid flow, because the model only imposes time-dependent curvature forces and employs geometric thresholds for the switch between power and recovery stroke, and vice versa. Figure 8 displays the beating pattern of an individual filament with an asymmetric power and recovery stroke. The hydrodynamically induced metachronal waves are visible in Fig. 8 (right).

## 6.2    Aggregation of Squirmers

The interactions between microswimmers depend on their relative orientation. Thereby, interactions of pushers and pullers in equivalent positions and orientations are equal in magnitude but opposite in sign, because of the opposite sign of their dipole strength $P$. Since swimmers typically meet at different relative positions and orientations and due to the stochastic motion of many interacting objects, scattering of microswimmers occurs, and cooperative swimming is the exception rather than
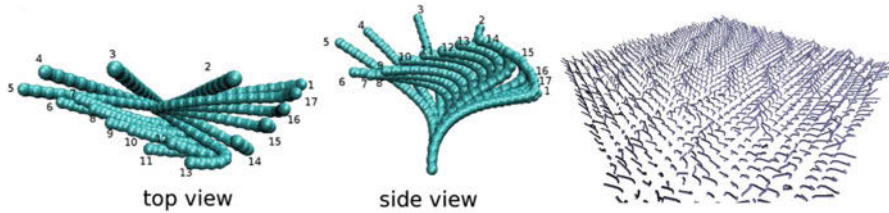
**Fig. 8** (Left) Top and side view of the beat pattern of the computational cilia model. Subsequent conformations are equally spaced in time. The fast, planar power stroke (frames 1−5) continues until a positive curvature threshold in the lower part of the cilium is reached. The cilium then switches to a slow, out-of-plane recovery stroke (frames 6−17), which ends when a negative curvature threshold is exceeded. (Right) Simulation snapshot of an array of $40 \times 40$ beating cilia. Cilia are placed on a square lattice. The metachronal wave is easily recognized by the lines of fully extended cilia during the power stroke. (From Elgeti and Gompper 2013)

the rule in dilute suspensions (Ishikawa et al. 2006; Llopis and Pagonabarraga 2010; Götze and Gompper 2010). This applies to spherical as well as spheroidal squirmers. However, hydrodynamic interactions with confining surfaces in a slit geometry stabilize the cooperative swimming of spheroidal pulling squirmers, which emphasizes the relevance of hydrodynamic interactions in confinement (Theers et al. 2016).

Self-propelled particles exhibit a strong tendency for clustering and phase separation. The origin of this behavior is the blockage of motion when several particles collide with each other (cf. Fig. 9). The particular (small) cluster would dissolve after a time corresponding to the reorientation time of a swimmer. Interactions and collisions with other particles are controlled by the density and propulsion velocity $v_0$. Hence, if other particles collide before the original cluster dissolved, the cluster grows. This behavior already occurs for active Brownian particles (Bialké et al. 2012; Buttinoni et al. 2013; Redner et al. 2013; Palacci et al. 2013; Marchetti et al. 2016; Bechinger et al. 2016), i.e., for self-propelled particles with steric but without hydrodynamic interactions. The separation into a dense solid (2D) (Bialké et al. 2012; Marchetti et al. 2016; Bechinger et al. 2016) or fluid (3D) (Wysocki et al. 2014) phase and a dilute gas phase is denoted as motility-induced phase separation (MIPS) (Cates and Tailleur 2015). Hydrodynamic interactions strongly modify the collective behavior.

The collective swimming patterns of spherical squirmers in 2D exhibit a strong dependence on the sign of the force dipole (far field). Moreover, hydrodynamic near-field effects play an important role. The phase behavior of neutral squirmers ($\beta = 0$) with only far-field interactions is similar to that of active Brownian particles without hydrodynamic interactions (Yoshinaga and Liverpool 2017). The additional contribution due to hydrodynamics is an enhanced reorientation of the squirmers, which suppresses phase separation. Squirmer ensembles in 2D without thermal motion exhibit clustering for pullers and pushers due to near-field hydrodynamic interactions. Neutral squirmers spontaneously develop polar order and collectively
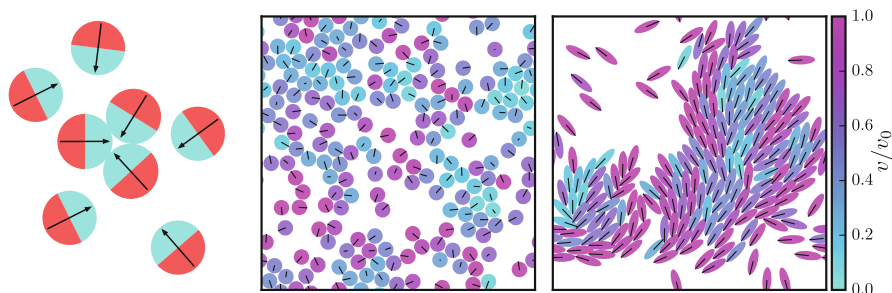
**Fig. 9** (Left) Aggregation and cluster formation of squirmers by blockage. (Middle) Fluid phase of spherical squirmers and (Right) cluster of spheroidal squirmers confined in a slit at the two-dimensional packing fraction $\phi^{2D} = 0.5$ and the Péclet number $Pe = 12$. The Péclet number is defined as $Pe = v_0/2b_z D_R^\perp$, where $b_z$ is the longer spheroid semiaxis and $D_R^\perp$ the rotational diffusion coefficient around the minor axis

move in a preferred direction (Kyoya et al. 2015; Yoshinaga and Liverpool 2017). Yet polar order is destroyed by thermal fluctuations. With fluctuations, all spherical squirmers exhibit cluster formation, however, with distinct characteristics. Here, cluster formation is most pronounced for pullers (Alarcón et al. 2017).

The anisotropic shape of a spheroidal squirmer enhances cluster formation compared to spherical swimmers (cf. Fig. 9) (Ginelli et al. 2010; Abkenar et al. 2013). This applies to ABPs as well as hydrodynamically coupled swimmers. Thereby, hydrodynamic interactions suppress motility-induced phase separation for spheres, whereas for spheroids hydrodynamic interactions enhance cluster formation in a slit geometry (Theers et al. 2018).

# 7    Conclusions

Hydrodynamic interactions are essential for active matter, specifically biological microswimmers. They are not only fundamental for the propulsion of microswimmers but also determine their behavior next to surfaces as well as the emergent collective dynamics and structures. Hydrodynamic interactions imply a very rich dynamics, which depends on the detailed swimming mechanism. For a fundamental understanding of the fluid-mediated interactions, consideration of the dominant multipole terms might suffice. However, the detailed collective properties depend also on the actual near-field flow. Hence, the full flow field has to be taken into account for a quantitative understanding of the local features of microswimmer aggregates. As a general conclusion, hydrodynamic interactions have to be taken into account for a qualitative and quantitative understanding of the emerging properties of active matter.

# References

Abkenar M, Marx K, Auth T, Gompper G (2013) Collective behavior of penetrable self-propelled rods in two dimensions. Phys Rev E 88:062314

Afzelius B (1976) A human syndrome caused by immotile cilia. Science 193:317

Alarcón F, Valeriani C, Pagonabarraga I (2017) Morphology of clusters of attractive dry and wet self-propelled spherical particle suspensions. Soft Matter 13:814

Bechinger C, Di Leonardo R, Löwen H, Reichhardt C, Volpe G, Volpe G (2016) Active particles in complex and crowded environments. Rev Mod Phys 88:045006

Berg HC (2003) The rotary motor of bacterial flagella. Annu Rev Biochem 72:19

Berke AP, Turner L, Berg HC, Lauga E (2008) Hydrodynamic attraction of swimming microorganisms by surfaces. Phys Rev Lett 101:038102

Bialké J, Speck T, Löwen H (2012) Crystallization in a dense suspension of self-propelled particles. Phys Rev Lett 108:168301

Blake JR (1971) A spherical envelope approach to ciliary propulsion. J Fluid Mech 46:199

Brumley DR, Polin M, Pedley TJ, Goldstein RE (2012) Hydrodynamic synchronization and metachronal waves on the surface of the colonial alga *Volvox carteri*. Phys Rev Lett 109:268102

Buttinoni I, Bialké J, Kümmel F, Löwen H, Bechinger C, Speck T (2013) Dynamical clustering and phase separation in suspensions of self-propelled colloidal particles. Phys Rev Lett 110:238301

Calladine CR (1975) Construction of bacterial flagella. Nature 255:121

Cartwright JHE, Piro O, Tuval I (2004) Fluid-dynamical basis of the embryonic development of left-right asymmetry in vertebrates. Proc Natl Acad Sci USA 101:7234

Cates ME, Tailleur J (2015) Motility-induced phase separation. Annu Rev Condens Matter Phys 6:219

Copeland MF, Weibel DB (2009) Bacterial swarming: a model system for studying dynamic self-assembly. Soft Matter 5:1174

Dhont JKG (1996) An introduction to dynamics of colloids. Elsevier, Amsterdam

Di Leonardo R, Dell'Arciprete D, Angelani L, Iebba V (2011) Swimming with an image. Phys Rev Lett 106:038101

Drescher K, Goldstein RE, Tuval I (2010) Fidelity of adaptive phototaxis. Proc Natl Acad Sci USA 107:11171

Drescher K, Dunkel J, Cisneros LH, Ganguly S, Goldstein RE (2011) Fluid dynamics and noise in bacterial cell-cell and cell-surface scattering. Proc Natl Acad Sci USA 10940:108

Elgeti J, Gompper G (2013) Emergence of metachronal waves in cilia arrays. Proc Natl Acad Sci USA 110:4470

Elgeti J, Gompper G (2016) Microswimmers near surfaces. Eur Phys J Spec Top 225:2333

Elgeti J, Winkler RG, Gompper G (2015) Physics of microswimmers—single particle motion and collective behavior: a review. Rep Prog Phys 78:056601

Ginelli F, Peruani F, Bär M, Chaté H (2010) Large-scale collective properties of self-propelled rods. Phys Rev Lett 104:184502

Götze IO, Gompper G (2010) Mesoscale simulations of hydrodynamic squirmer interactions. Phys Rev E 82:041921

Gray J, Hancock GJ (1955) The propulsion of sea-urchin spermatozoa. J Exp Biol 32:802

Heinrichsen J (1978) Bacterial surface translocation: a survey and a classification. Bacteriol Rev 36:478

Hu J, Wysocki A, Winkler RG, Gompper G (2015a) Physical sensing of surface properties by microswimmers – directing bacterial motion via wall slip. Sci Rep 5:9586

Hu J, Yang M, Gompper G, Winkler RG (2015b) Modelling the mechanics and hydrodynamics of swimming *E. coli*. Soft Matter 11:7843

Ishikawa T (2009) Suspension biomechanics of swimming microbes. J R Soc Interface 6:815

Ishikawa T, Simmonds MP, Pedley TJ (2006) Hydrodynamic interaction of two swimming model micro-organisms. J Fluid Mech 568:119

Kearns DB (2010) A field guide to bacterial swarming motility. Nat Rev Microbiol 8:634

Keller SR, Wu TY (1977) A porous prolate-spheroidal model for ciliated micro-organisms. J Fluid Mech 80:259

Kim S, Karrila SJ (1991) Microhydrodynamics: principles and selected applications. Butterworth-Heinemann, Boston

Kyoya K, Matsunaga D, Imai Y, Omori T, Ishikawa T (2015) Shape matters: near-field fluid mechanics dominate the collective motions of ellipsoidal squirmers. Phys Rev E 92:063027

Lauga E, Powers TR (2009) The hydrodynamics of swimming microorganisms. Rep Prog Phys 72:096601

Lauga E, DiLuzio WR, Whitesides GM, Stone HA (2006) Swimming in circles: motion of bacteria near solid boundaries. Biophys J 90:400

Lighthill MJ (1952) On the squirming motion of nearly spherical deformable bodies through liquids at very small Reynolds numbers. Comm Pure Appl Math 5:109

Llopis I, Pagonabarraga I (2010) Hydrodynamic interactions in squirmer motion: swimming with a neighbour and close to a wall. J Non-Newtonian Fluid Mech 165:946

López HM, Gachelin J, Douarche C, Auradou H, Clément E (2015) Turning bacteria suspensions into superfluids. Phys Rev Lett 115:028301

Macnab RM (1977) Bacterial flagella rotating in bundles: a study in helical geometry. Proc Natl Acad Sci USA 74:221

Marchetti MC, Fily Y, Henkes S, Patch A, Yllanes D (2016) Minimal model of active colloids highlights the role of mechanical interactions in controlling the emergent behavior of active matter. Curr Opin Colloid Interface Sci 21:34

Moore HDM, Taggart DA (1995) Sperm pairing in the opossum increases the efficiency of sperm movement in a viscous environment. Biol Reprod 52:947

Palacci J, Sacanna S, Steinberg AP, Pine DJ, Chaikin PM (2013) Living crystals of light-activated colloidal surfers. Science 339:936

Platzer J, Sterr W, Hausmann M, Schmitt R (1997) Three genes of a motility operon and their role in flagellar rotary speed variation in *Rhizobium meliloti*. J Bacteriol 179:6391

Popkin G (2016) The physics of life. Nature 529:16

Pozrikidis C (1992) Boundary integral and singularity methods for linearized viscous flow. Cambridge University Press, Cambridge

Purcell EM (1977) Life at low Reynolds number. Am J Phys 45:3

Qian B, Jiang H, Gagnon DA, Breuer KS, Powers TR (2009) Minimal model for synchronization induced by hydrodynamic interactions. Phys Rev E 80:061919

Redner GS, Hagan MF, Baskaran A (2013) Structure and dynamics of a phase-separating active colloidal fluid. Phys Rev Lett 110:055701

Reichert M, Stark H (2005) Synchronization of rotating helices by hydrodynamic interactions. Eur Phys J E 17:493

Reigh SY, Winkler RG, Gompper G (2012) Synchronization and bundling of anchored bacterial flagella. Soft Matter 8:4363

Saintillan D (2010) The dilute rheology of swimming suspensions: a simple kinetic model. Exp Mech 50:1275

Sivinski J (1984) Sperm in competition. In: Smith RL (ed) Sperm competition and the evolution of animal mating systems. Academic, Orlando, p 174

Sleigh MA (1962) The Biology of Cilia and Flagella. Pergamon Press, Oxford

Spagnolie SE, Lauga E (2012) Hydrodynamics of self-propulsion near a boundary: predictions and accuracy of far-field approximations. J Fluid Mech 700:105

Theers M, Westphal E, Gompper G, Winkler RG (2016) Modeling a spheroidal microswimmer and cooperative swimming in a narrow slit. Soft Matter 12:7372

Theers M, Westphal E, Qi K, Winkler RG, Gompper G (2018) Clustering of microswimmers: interplay of shape and hydrodynamics. Soft Matter 14:8590–8603

Tirado MM, Martínez CL, de la Torre JG (1984) Comparison of theories for the translational and rotational diffusion coefficients of rod-like macromolecules. Application to short DNA fragments. J Chem Phys 81:2047

Vicsek T, Zafeiris A (2012) Collective motion. Phys Rep 517:71

Wang Q, Pan J, Snell WJ (2006) Intraflagellar transport particles participate directly in cilium-generated signaling in *Chlamydomonas*. Cell 125:549

Wensink HH, Dunkel J, Heidenreich S, Drescher K, Goldstein RE, Löwen H, Yeomans JM (2012) Meso-scale turbulence in living fluids. Proc Natl Acad Sci USA 109:14308

Winkler RG (2016) Low Reynolds number hydrodynamics and mesoscale simulations. Eur Phys J Spec Top 225:2079

Wysocki A, Winkler RG, Gompper G (2014) Cooperative motion of active Brownian spheres in three-dimensional dense suspensions. EPL 105:48004

Yeomans JM, Pushkin DO, Shum H (2014) An introduction to the hydrodynamics of swimming microorganisms. Eur Phys J Spec Top 223:1771

Yoshinaga N, Liverpool TB (2017) Hydrodynamic interactions in dense active suspensions: from polar order to dynamical clusters. Phys Rev E 96:020603

Zöttl A, Stark H (2014) Hydrodynamics determines collective motion and phase behavior of active colloids in quasi-two-dimensional confinement. Phys Rev Lett 112:118101

Zöttl A, Stark H (2016) Emergent behavior in active colloids. J Phys Condens Matter 28:253

# Part X
# Crystal Plasticity: From the Atomic Scale to the Macroscale

# Recent Advances in Crystal Plasticity Modeling

# 69

Wei Cai and Somnath Ghosh

## Contents

**Abstract**

Chapters in this section introduce recent advances in computational models and numerical methods for establishing a physics-based understanding of crystal plasticity, starting from atomistic models of dislocations, to discrete models of dislocation line networks, to continuum descriptions of microstructure evolution.

## 1    Purpose of Physics-Based Modeling

The plastic deformation of crystalline materials (especially metals and alloys) is a fundamental process that has been critically important for the technologies of mankind for thousands of years. Understanding the process of crystal plasticity is an essential step for addressing many of today's technological needs, by enabling

W. Cai (✉)
Departments of Mechanical Engineering, Stanford University, Stanford, CA, USA
e-mail: caiwei@stanford.edu

S. Ghosh
JHU Center for Integrated Structure-Materials Modeling and Simulation (CISMMS), Air Force Center of Excellence on Integrated Materials Modeling (CEIMM), US Association of Computational Mechanics (USACM), Department of Civil Engineering, Departments of Mechanical Engineering and Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA
e-mail: sghosh20@jhu.edu

precise control of the deformation process and design of new materials with tailored properties. The application areas where such an understanding has an impact include metal forming, machining, additive manufacturing, and material design for high-temperature components (such as turbine blades), to name a few. Crystal plasticity also plays an important role in other processes such as ductile fracture and fatigue.

Our initial understanding of crystal plasticity came from experiments (such as tensile and torsion tests) and phenomenological continuum theories (such as $J_2$ plasticity theory) that are consistent with observations and symmetry requirements. With the latest advances of experimental characterization techniques (e.g., accurate 3D representation of grain structure of a polycrystal) and computational tools (e.g., parallel processors, GPUs), physics-based models of crystal plasticity that account for granular, subgranular, and atomic-level processes are becoming ever more realistic and predictive. Bridges between such models at different scales are starting to be established. A well-connected multiscale framework for crystal plasticity may be established in the near future. Such a framework is needed to realize the full potential of physics-based predictive models in order to make an impact on application areas in manufacturing and materials design.

In this section, we present ten chapters that introduce recent advances on physics-based crystal plasticity models at different scales, starting from atomistic models of the dislocation core, to discrete models of dislocation line networks, to continuum descriptions of subgranular and intergranular microstructure evolution. In addition to providing new researchers entering this field an introduction to the modeling techniques they may want to learn, we hope that by looking at the latest advances at crystal plasticity models at all scales, this section also helps to identify the opportunities and the gaps in the multiscale framework that may also be useful to an experienced researcher.

## 2    Multiscale Challenge of Crystal Plasticity

Under most conditions (e.g., temperature below $\sim 1/3$ of the melting point, strain rate below $\sim 10^5\,\text{s}^{-1}$), the plastic deformations of metals and alloys are predominantly carried out by the motion of dislocations, which are line defects in the crystal (Hull and Bacon 2011; Cai and Nix 2016; Anderson et al. 2017). This is why dislocations feature prominently in physics-based models of crystal plasticity. At the lower limit of the length scales of interest here, the electronic structure in the atomistic region around the dislocation core determines the energetics and mobility of the dislocation. At the upper limit of the length scale, engineering materials are usually available in the form of a polycrystal, consisting of many grains whose size and orientation distributions (i.e., texture) influence the overall response of the material. Establishing a quantitative understanding of crystal plasticity is challenging because of the wide range of length scales (and an even wider range of time scales) between these two limits, as shown in Fig. 1. No single theoretical/computational model can cover the entire range of the relevant length
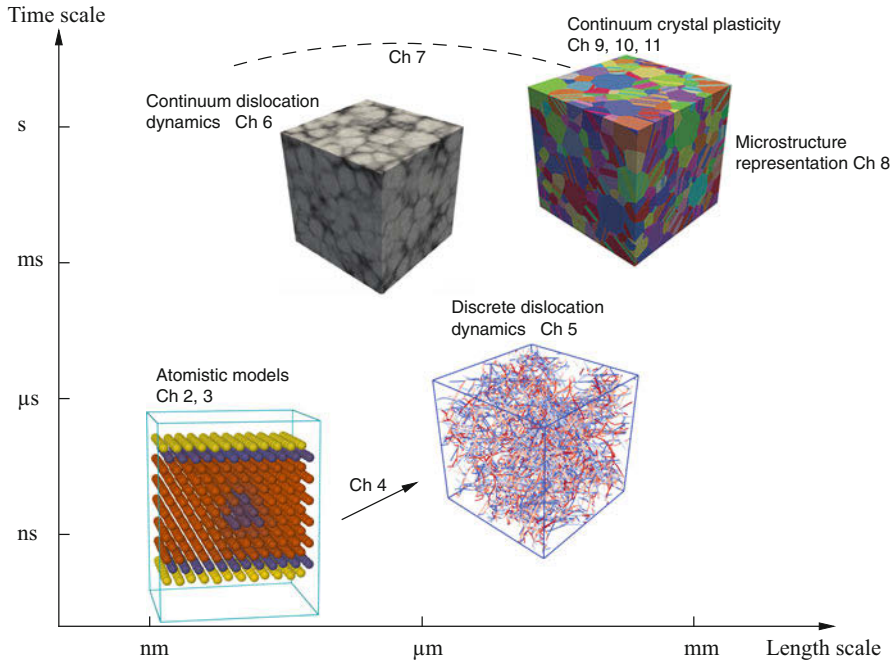
**Fig. 1** Schematics of crystal plasticity models at different length and time scales and the relationship between different chapters in this section

and time scales, which is why different models have been developed and ultimately need to be connected to reach a comprehensive understanding of crystal plasticity.

The chapters in this section progress generally from smaller to larger length and time scales as shown in Fig. 1. In ▶ Chap. 70, "Ab Initio Models of Dislocations" (by E. Clouet), we start with ab initio models of the dislocation core. These are atomistic models that also account for the quantum mechanics of electrons based on the density functional theory. In the multiscale hierarchy here, these models are the most accurate and contain the fewest of empirical parameters, but their high computational cost limits them to very small material volumes typically containing at most a few hundred atoms. As a result, boundary conditions are important considerations for the successful applications of such models. A significant emphasis of Chap. 2 is on periodic boundary conditions (PBC). Chapter 2 discusses how key dislocation properties such as core structure, core energy, and Peierls stress (i.e., the critical stress required to move a dislocation at zero temperature) can be predicted by ab initio models under PBC. These properties can be used to construct functions that describe forces on and mobility of dislocations that can be used in higher-scale dislocation dynamics (DD) models (Chap. 4).

In ▶ Chap. 71, "Modeling the Thermally Activated Mobility of Dislocations at the Atomic Scale" (by L. Proville and D. Rodney), we introduce atomistic models of dislocations in which interatomic interactions are described by empirical

potentials. The simplification (i.e., removal of quantum mechanics of the electrons) allows such models to handle much larger material volumes, containing millions of atoms. This allows molecular dynamics (MD) simulations of the motion of fairly long dislocation lines (e.g., 100 Burgers vectors in length) over time scales of nanoseconds, so that mobility (a key input to DD models, Chap. 5) of fast moving dislocations (e.g., those in face-centered cubic metals) at finite temperature can be directly extracted. However, in materials with high Peierls stress (e.g., screw dislocations in body-centered cubic metals), dislocation motion can be thermally activated, and dislocation velocity would be too slow to be observable by direct MD simulations. Chapter 3 presents computational methods based on the transition state theory to determine the dislocation mobility in these circumstances. It also discusses how to incorporate quantum corrections (due to the zero-point motion of the nuclei) to the predicted dislocation mobility.

In ▶ Chap. 72, "Dislocation Analysis Tool for Atomistic Simulations" (by A. Stukowski), we introduce methods which convert atomistic descriptions of dislocations into line descriptions in the discrete dislocation dynamics (DDD) model (which is the subject of Chap. 5). Doing so leads to a tremendous reduction in the number of degrees of freedom, so that DDD models can deal with much larger material volumes, e.g., with sizes exceeding $10\,\mu$m. Chapter 3 introduces the dislocation extraction algorithm (DXA), which automatically identifies dislocations in an atomistic simulation snapshot and determines their Burgers vectors. This capability not only facilitates the visualization of atomistic simulations of crystal plasticity but also opens the exciting possibility of one-to-one comparisons between MD and DDD simulations of the same dislocation ensemble. This would allow higher-scale models (such as DDD) to be benchmarked against more fundamental, lower-scale models (such as MD), thus strengthening the multiscale hierarchy of crystal plasticity.

In ▶ Chap. 73, "Line Dislocation Dynamics Simulations with Complex Physics" (by R. B. Sills and S. Aubry), we present recent advances in discrete dislocation dynamics (DDD) models, which take core energy and mobility inputs from atomistic models and predict single crystal stress-strain relations that can be used in continuum plasticity models at even higher scales. Chapter 5 focuses on the recent developments of DDD models on the more "complex" physical aspects of dislocation processes relevant to crystal plasticity. These include, for example, elastic anisotropy, nonlinear mobility laws, and interactions with solutes, vacancies, precipitates, and grain boundaries.

While DDD models have the potential to link fundamental dislocation processes with plastic deformation behavior of single crystals and to predict spontaneous pattern formation in the dislocation microstructure, currently DDD models are still very much limited in the amount of strain they can reach (e.g., about a few percent) under work-hardening conditions. It is possible that directly connecting DDD models with continuum crystal plasticity models may still be impractical (or uneconomical) in the near future, and another model may be needed to bridge the two.

In ▶ Chap. 74, "Continuum Dislocation Dynamics: Classical Theory and Contemporary Models" (by A. El-Azab and G. Po), we present the continuum dislocation dynamics (CDD) model that is motivated by this need. The degrees of freedom in the CDD model are (smooth) fields (e.g., dislocation densities) represented on a volumetric grid. The grid sizes (e.g., 30 nm) are typically larger than minimum segment lengths in DDD models. The explicit network structure in the DDD model is "smeared out" and replaced by density variations in space, so that the CDD model can be considered as the result of "coarse-graining" the DDD model. The main challenge in CDD models is to decide which degrees of freedom (i.e., microstructural information beyond total density) should be kept after the coarse-graining and what their equations of motion should be. At present, multiple CDD formulations exist in the literature, and a number of them are discussed in Chap. 6. It is exciting to see that some of the 3D CDD formulations are already able to generate dislocation density patterns that are similar to experimental observations. Establishing a closer comparison between DDD and CDD models (e.g., through a systematic process of coarse-graining) appears to be a promising and important task for establishing a robust multiscale hierarchy of crystal plasticity models.

The next few chapters from 7 to 11 have a major focus on higher length scales, those corresponding to single and polycrystals in the microstructure. They propose alternative approaches for solving crystal plasticity problems in polycrystalline microstructures, overcoming some of the bottlenecks that jeopardize their use in complex applications. While a quantitative connection with the more fundamental, dislocation-based models is still lacking, impressive progress has been made over the last decade in crystal plasticity models in terms of numerical efficiency and the ability to model realistic microstructures.

In ▶ Chap. 75, "Connecting Lower and Higher Scales in Crystal Plasticity Modeling" (by D. L. McDowell) a generalized description of crystalline plasticity is proposed to encompass a spectrum of models that address various phenomena associated with dislocation evolution in crystals across a range of length and time scales. The evolution processes involve dislocation nucleation, generation, migration, interaction, trapping, and annihilation in crystals and polycrystals. Starting from coarse-grained atomistics, the models described include microscopic phase-field models, field dislocation models, discrete dislocation dynamics, statistical continuum dislocation models, mesoscale generalized continuum models categorized as gradients, micropolar or micromorphic types, all the way up to local continuum plasticity of polycrystalline aggregates. Furthermore the concepts of concurrent and hierarchical multiscale modeling in spatial and temporal domains are discussed. The chapter concludes with a summary of the challenging and long-standing gaps in hierarchical modeling of crystalline plasticity models, including quantification of coarse-graining and model reduction errors and uncertainty quantification of individual models, domain decomposition, as well as various two-scale transitions.

In ▶ Chap. 76, "Developing Virtual Microstructures and Statistically Equivalent Representative Volume Elements for Polycrystalline Materials" (by S. Ghosh and M. Groeber), the development of computational methods for simulating statistically

equivalent virtual microstructures of materials with complex heterogeneities is discussed. Specifically, the methods focus on polycrystalline metallic materials containing localized features such as annealing twins, particulates or precipitates, and subgrain-scale precipitates. The materials discussed include alloys of titanium, magnesium, and aluminum, as well as nickel-based superalloys. Data from image analysis and characterization are used to construct 3D statistical distribution and correlation functions. This serves as input to the virtual microstructure generation process. The methods accommodate computational approaches that infer 3D statistical descriptors and functions from 2D observations and surface data from stereology and other optimization-based projection techniques. The chapter introduces the software package DREAM.3D that is widely used for generating virtual microstructures of polycrystalline materials. A host of newly developed methods for incorporating twins, particles, and subgrain-scale phases in polycrystalline microstructures are further discussed. Finally the concept of the microstructure-based SERVE or M-SERVE, in which morphological and crystallographic characteristics of the microstructure are determinants of the statistically equivalent representative volume element, is introduced.

In ▸ Chap. 77, "Polycrystal Plasticity Models Based on Green's Functions: Mean-Field Self-Consistent and Full-Field Fast Fourier Transform Formulations" (by R. Lebensohn), two Green's function-based crystal plasticity methodologies for predicting microstructure-property relations in polycrystalline aggregates are discussed: The mean-field viscoplastic self-consistent (VPSC) method and the full-field viscoplastic fast Fourier transform method (VPFFT). These formulations are based on the Green's function solution of the PDEs governing the micromechanical response of polycrystals. The heterogeneity is related to the inherent directional properties of grains along different crystal directions and crystallographic orientations of each grain in the aggregate. The VPSC formulation uses a linearization assumption of the behavior of single crystals and utilizes Fourier transforms in an ellipsoidal domain of the Green's function for an infinite medium with homogenous properties. The VPFFT method evaluates convolution integrals over periodic unit cells using the Green's function for a periodic medium and a polarization term containing all the information on the heterogeneity and nonlinearity of the crystalline material behavior. The chapter delineates the similarities and differences of both the crystal plasticity formulations and cross-validates their predictions.

In ▸ Chap. 78, "Computationally Efficient Crystal Plasticity Simulations Using Spectral Databases" (by S. R. Kalidindi, A. Gupta and E. Popova), a spectral approach to crystal plasticity solutions is presented. The spectral database approach is able to significantly enhance the computational efficiency of the crystal plasticity framework. Important variables from crystal plasticity computations are accessed directly from precomputed spectral databases based on discrete Fourier transform (DFT). The chapter discusses the DFT database generation process for deviatoric stresses, lattice spins, and strain-hardening rates using Taylor-type crystal plasticity models. Studies are performed on Taylor-type simulations for FCC and BCC materials, yield surface evaluation, property closures, as well as forming limit diagrams. They show considerable computational speedup due to reduced level of

computing with this approach. The extension of the spectral database framework to crystal plasticity FEM simulations is also discussed. The advantages of using the spectral crystal plasticity method with graphics processing unit (GPU) computing are predicted.

In ▶ Chap. 79, "Advances in Computational Mechanics to Address Challenges in Crystal Plasticity FEM" (by S. Ghosh), novel methods and algorithms in computational mechanics are discussed for facilitating robust and efficient crystal plasticity finite element (CPFE) modeling of deformation and failure in metals and alloys. Image-based CPFE models incorporate characteristic microstructural features, as well as underlying physical mechanisms. A number of challenges arise when using CPFE models for modeling deformation mechanisms in complex microstructures, especially those involving phenomena like localization, twinning, crack propagation, fatigue, etc. This chapter examines three challenges that commonly persist with conventional CPFEM and offers remedies for overcoming them. The three problems include element stabilization to overcome plastic incompressibility due to volumetric locking of tetrahedral elements in CPFE analyses, time-domain subcycling to account for disparate deformation rates in modeling localization phenomena like discrete twin evolution in CPFEM, and wavelet-induced adaptive hierarchical CPFEM for enhanced efficiency. The chapter ends with a discussion of multi-time scaling issues in fatigue problems.

## References

Anderson PM, Hirth JP, Lothe J (2017) Theory of dislocations, 3rd edn. Cambridge University Press, Cambridge

Cai W, Nix WD (2016) Imperfections in crystalline solids. Cambridge University Press, Cambridge

Hull D, Bacon DJ (2011) Introduction to dislocations, 5th edn. Elsevier, Oxford

# Ab Initio Models of Dislocations

# 70

## Emmanuel Clouet

## Contents

**Abstract**

This chapter reviews the different methodological aspects of the ab initio modeling of dislocations. Such simulations are now frequently used to study the dislocation core, i.e., the region in the immediate vicinity of the line defect where the crystal is so strongly distorted that an atomic description is needed. This core region controls some dislocation fundamental properties, like their ability to glide in different crystallographic planes. Ab initio calculations based on the density functional theory offer a predictive way to model this core region. Because

E. Clouet (✉)
DEN-Service de Recherches de Métallurgie Physique, CEA, Gif-sur-Yvette, France
e-mail: emmanuel.clouet@cea.fr

dislocations break the periodicity of the crystal and induce long-range elastic fields, several specific approaches relying on different boundary conditions have been developed to allow for the atomistic modeling of these defects in simulation cells having a size compatible with ab initio calculations. We describe these different approaches which can be used to study dislocations with ab initio calculations and introduce the different analyses which are currently performed to characterize the core structure, before discussing how meaningful energy properties can be extracted from such simulations.

# 1    Introduction

Dislocations are line defects which control the development of the plastic deformation in crystals. These defects induce a long-range stress field, which is well described by elasticity, and dislocation elasticity theory offers a powerful framework to model dislocations and their interaction with their surrounding environment (Hirth and Lothe 1982; Bacon et al. 1980). But some of their fundamental properties, like their glide plane and their mobility, highly depend on their core, i.e., the region in the immediate vicinity of the defect where the perturbation of the crystal is too important to be described by elasticity. The modeling of this core region necessitates an atomic description, and atomistic simulations have thus become a valuable tool to study dislocation properties. Among such simulations, ab initio calculations based on the density functional theory (DFT), as they rely on an electronic description of the atomic bonding, appear as the most accurate and predictive. But as these calculations are still limited in the size of the system they can handle, typically at most a few hundred atoms, the ab initio modeling of dislocations need special attention. Specific methodologies have been therefore developed to study dislocation core properties with ab initio calculations. The purpose of this chapter is to review the different modeling approaches for the ab initio study of dislocations, starting from a quick overview of DFT formalism, before describing more thoroughly boundary conditions specific to dislocation models, then the analysis of the atomic structure in the dislocation core and finally the extraction of meaningful energy properties. Beyond the examples illustrated in this chapter, results which have been obtained from such ab initio studies for the dislocation core properties in different metals and semiconductors can be found in the recent review of Rodney et al. (2017).

# 2    Ab Initio Calculations

Ab initio calculations describe the bonding between atoms, thanks to the resolutions of the Schrödinger equation for the electrons of the system. These are first-principles approaches as they do not use any experimental data and allows the modeling of atomic interaction only from the atomic number and other fundamental quantities. Compared to empirical interatomic potentials, such approaches are completely

transferable, without any parameterization depending on the environment under study, but at the expense of a much higher CPU time. Although ab initio in nature and usually very accurate, these approaches nevertheless rely on different approximations, the validity of which needs generally to be assessed.

The most fundamental approximation is the Born-Oppenheimer approximation. As atom nuclei have a much higher mass than electrons, one can assume that the electrons are always equilibrated with respect to the positions of the nuclei which are considered as immobile. The resolution of the Schrödinger equation for the electrons therefore leads to the energy of the system as a function of the atomic positions. Knowing this function and also its first derivatives, i.e., the atomic forces, standard algorithms of atomic simulations can then be used. For the ab initio modeling of dislocations, this is usually restricted to molecular statics, including energy barrier calculations, because of the high CPU burden of the energy and force calculation.

Most ab initio calculations of dislocations are relying on the density functional theory (DFT). This makes use of the Hohenberg and Kohn (1964) theorem showing that the ground-state energy is the minimum of a functional depending only on the electronic density. This dramatically simplifies the problem as the electronic density depends only on the position, whereas the many-electron wave function entering Schrödinger equation is a function depending on the $3N$ electron coordinates, with $N$ the number of electrons in the system. The Kohn and Sham (1965) approach allows then a practical implementation, where the Schrödinger equation is solved for an equivalent system of noninteracting electrons. This necessitates the definition of an unknown contribution to the Hamiltonian, the exchange and correlation potential. Most of dislocation calculations are performed with the local density (LDA) or the generalized gradient (GGA) approximations, assuming that this contribution depends only locally on the electronic density or also its gradient.

For dislocation calculations, it is enough to consider that only the electrons of the outer shells participate to the atomic bonding. Electrons of the inner shells are not sensitive to the atom environment and can be assumed to have the same ground state as for the isolated atom. Kohn-Sham equations are then solved only for valence electrons. One can further reduce the CPU overhead by replacing with a pseudopotential the interaction potential of the valence electrons with the ionic core. This pseudopotential aims to reduce the strong oscillations of the electronic wave functions close to the dislocation core, because the description of these oscillations necessitates a large basis set, while still leading to the correct wave functions outside this core region. Different pseudoization schemes, norm-conserving or ultrasoft pseudopotentials as well as the projected augmented wave (PAW) method, are available.

Ab initio codes used for dislocations are relying on Born-von Karman periodic boundary conditions to model the solid, whatever the boundary conditions used to incorporate a dislocation in the simulation cell (Sect. 3). Electronic wave functions are thus a superposition of Bloch waves with wavevectors spanning the first Brillouin zone. Integration in the reciprocal space is performed on a regular grid sampling the first Brillouin zone, using smearing functions to broaden the electronic

density of states. Different basis sets can be used to describe the Bloch waves, with plane waves being the most popular choice for dislocations.

Ab initio approaches devoted to the study of dislocations are thus not specific: they are making use of standard DFT implementations which are now current modeling tools in solid-state physics. Feature specific to dislocation modeling, as described in the next section, is the necessity to use a supercell large enough to let the dislocation core adopt its fully relaxed configuration, with boundary conditions compatible with the long-range distortion induced by the defect. A high accuracy is also generally needed for such calculations as the energy variations involved by the dislocation core are usually small. For instance, the Peierls energy barrier opposing the glide of $1/2 \langle 111 \rangle$ screw dislocations in BCC transition metals does not exceed $100 \, \text{meV}/b$, where $b$, the norm of the Burgers vector, corresponds to the height of the simulation cell necessary to model such a dislocation.

## 3    Boundary Conditions

The ab initio modeling of dislocations needs special care in the way the boundary conditions are handled. First, a dislocation creates a long-range elastic field which needs to be taken into account. Second, it is not possible to include a single dislocation in a simulation box with full periodic boundary conditions which usually constitute the paradigm in the modeling of bulk materials: the dislocation opens a displacement discontinuity, and another defect is needed to close the discontinuity and allow for periodicity. As a result, different boundary conditions compatible with ab initio calculations have been developed to model dislocations.

All approaches enforce periodicity in the direction of the dislocation line. In pure metals, one usually uses the shortest periodicity vector to define the dimension of the simulation cell in this direction, thus modeling an infinite straight dislocation. But this size needs to be increased if one wants to introduce a solute atom on the dislocation line, so as to minimize the interaction of the solute atom with its periodic images and truly study the interaction of the dislocation with a single foreign atom. A larger size is also needed to model a kinked dislocation. This is usually possible only in covalent crystals where the atomic bonds are highly directional, leading to abrupt kinks experiencing a non-negligible energy barrier when migrating along the dislocation line. In metallic systems with less directional atomic bonding, kinks are usually spread over a larger distance and are highly mobile, making it hard to stabilize them in a simulation cell whose size is compatible with ab initio calculations.

### 3.1    Cluster Approach

The easiest way to model a dislocation is to use an infinite cylinder whose axis coincides with the dislocation line. Periodicity is enforced only along the dislocation line. The dislocation is created by displacing all atoms according to the Volterra

**Fig. 1** Boundary conditions used to model an isolated straight dislocation in the cluster approach. The outer boundary is either (**a**) rigid or (**b**) flexible and controlled by lattice Green's functions or by coupling with an empirical potential
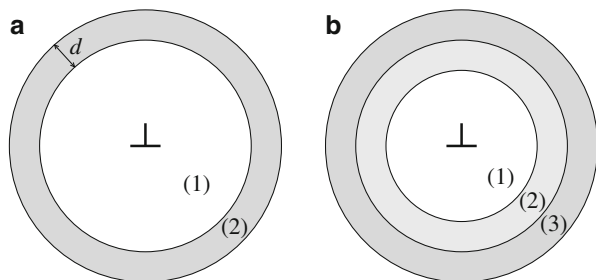
solution given by anisotropic elasticity theory for the dislocation displacement field (Stroh 1958, 1962). Atoms at the cylinder surface (region 2 in Fig. 1a) are kept fixed in their initial positions, and only atoms inside the cylinder are relaxed. One thus models an isolated dislocation in an infinite continuum.

But this modeling approach has severe drawbacks. The elastic solution used to fix the atoms at the boundary is only approximate as it relies on linear elasticity, thus neglecting crystal anharmonicity which can be strong close to the dislocation line. Moreover, the Volterra elastic solution, used to fix the atoms at the boundary, only corresponds to the long-range elastic field of the dislocation. Close to the dislocation line some additional contributions, the dislocation core field, need to be accounted for (Eshelby et al. 1953). A spreading of the dislocation core can be the reason for the existence of such a core field, but even dislocations with a compact core, like ⟨111⟩ screw dislocations in BCC metals, possess a non-negligible core field. Although this core field decays more rapidly than the Volterra elastic field, the size of the simulation boxes that can be handled by ab initio calculations is never large enough to neglect it. The rigid boundary conditions do not allow the correct development of this core field and thus perturb the relaxation of the dislocation core.

The fixed atomic positions imposed at the boundary also prevent use of this method to determine the lattice friction opposing dislocation motion. If the dislocation moves during the simulation, this boundary condition will not be compatible anymore with the new dislocation position. This induces indeed a back-stress opposing the dislocation motion. As a result, any simulation relying on this boundary condition will overestimate the dislocation Peierls stress, which is the minimum stress necessary to move the dislocation at 0 K.

## 3.2    Flexible Boundary Conditions

To remove the artifacts induced by the rigid boundary conditions, dislocation modeling with flexible boundary conditions has been developed. The proposed method relies either on the use of the lattice Green's function (Sinclair et al. 1978; Woodward 2005) or on the coupling with an empirical potential (Liu et al. 2007; Chen et al. 2008).

The lattice Green's function $G_{ij}(\mathbf{r})$ expresses, in the crystal harmonic approximation, the displacement $\mathbf{u}$ induced on an atom in position $\mathbf{r}$ by a force $\mathbf{F}$ acting on an atom at origin (We use the Einstein implicit summation convention on repeated indexes appearing in all expressions.):

$$u_i(\mathbf{r}) = G_{ij}(\mathbf{r})F_j. \tag{1}$$

This lattice Green's function can be obtained by inversion of the force-constant matrices of the perfect crystal (Yasi et al. 2012; Tan and Trinkle 2016) or can be tabulated from direct calculations in a perfect lattice (Sinclair et al. 1978; Rao et al. 1998). In the long-range limit, $G_{ij}(\mathbf{r})$ converges to the elastic Green's function given by anisotropic elasticity theory.

Flexible boundary conditions based on lattice Green's functions still make use of a cylinder geometry to model a single dislocation, but three zones are now defined (Fig. 1b). Atoms in the inner zone (1) are relaxed with the ab initio code until the forces acting on them are smaller than a fixed threshold, while atoms in zones (2) and (3) are kept fixed. At the end of this step, atomic forces have appeared in zone (2), because the dislocation elastic field deviates from the Volterra solution used as an initial guess. The lattice Green's function is then used to displace atoms in all three zones according to Eq. 1 using all atomic forces in zone (2). This leads to the cancelation of forces in zone (2) but makes new forces appear in zone (1). The procedure is thus iterated until all forces in zones (1) and (2) are null. This self-consistent cycle is necessary because the lattice Green's function of the perfect crystal only approximates the linear response of the dislocated crystal. Atoms in zone (3) serve as a buffer to prevent any perturbation by the external boundary of forces building in zone (2). As shown by Segall et al. (2003), this buffer region may need to be quite large in metals to obtain negligible perturbations in the inner regions. This can be minimized by removing the surfaces delineating zone (3) and using periodic boundary conditions in all directions. Interface defects are then present at the boundary between two periodic simulations cells. But these defects lead to a smaller perturbation of the electronic density than the vacuum layer of the surfaces (Woodward 2005). One thus perfectly models an isolated dislocation in an infinite crystal taking full account of the dislocation core field. It is possible to study dislocation cores with a reduced number of atoms in the simulation cell, a size usually compatible with ab initio calculations.

A similar approach relies on the coupling of the ab initio calculations with an empirical potential (Liu et al. 2007; Chen et al. 2008). The simulation cell is still divided in three regions (Fig. 1b). Ab initio calculations are performed only for a smaller simulation cell corresponding to regions (1) and (2). Atoms in regions (2) and (3) are relaxed according to the forces calculated with the empirical potential, whereas atoms in region (1) are relaxed according to ab initio forces plus a correction to withdraw the perturbation caused by the external boundary of the ab initio box. The buffer region (2) has been added to the original approach (Choly et al. 2005) to minimize this correction by protecting atoms from the external boundary. To operate, this method needs therefore an empirical potential which

perfectly reproduces the lattice parameters given by ab initio calculations, which can generally be done by rescaling the distances. Besides, the potential has also to match as best as possible the ab initio linear response, i.e., at least the elastic constants and, ideally, the whole phonon spectrum.

As it will be discussed in the last section, the main drawback of this ab initio dislocation model using flexible boundaries arises from the difficulty of extracting dislocation energy. The problem may be actually less sensitive with the second approach relying on a coupling with an empirical potential where an energy formulation exists. In this case, one can obtain a reasonable estimation of the dislocation energy provided the potential gives an accurate description of the boundary energy compared to the ab initio calculations. While these flexible boundaries truly allow the modeling of an isolated dislocation, thus predicting its core structure and its evolution under an applied stress without any a priori artifact induced by the small size of the simulation cell inherent to ab initio calculations, the approach is still under active development to also provide information on the dislocation energy.

## 3.3    Periodic Boundary Conditions

To get rid of the external boundary and to use periodic boundary conditions in all three directions without the introduction of a defective interface, one needs to introduce in the simulation cell a dislocation dipole, i.e., two dislocations with opposite Burgers vectors. One thus models a 2D periodic array of dislocations (Fig. 2).

Several arrangements of dislocation arrays can be though off, but they are not all equivalent. Among all of them, the ones which are quadrupolar display strong advantages. A periodic array is quadrupolar, if the vector $\mathbf{d}$ linking the two dislocations of opposite signs is equal to $1/2$ $(\mathbf{u}_1 + \mathbf{u}_2)$, where $\mathbf{u}_1$ and $\mathbf{u}_2$ are the periodicity vectors of the simulation cell (Fig. 2). This ensures that every dislocation is a symmetry center of the array: fixing, as a convention, the origin at a dislocation center, if a dislocation $\mathbf{b}$ is located at the position $\mathbf{r}$, there will also be a dislocation $\mathbf{b}$ in $-\mathbf{r}$. The stress created by these two dislocations will cancel to first order at the origin, thanks to the symmetry property of the Volterra elastic field. ($\sigma_V(-\mathbf{r}) = -\sigma_V(\mathbf{r})$ with $\sigma_V$ the Volterra stress field of a single dislocation.) As a consequence, this quadrupolar periodic array minimizes the elastic interaction between the dislocations and hence the Peach-Koehler force acting on each dislocation because of the image dislocations associated with periodic boundaries. It is the best-suited periodic array to extract dislocation core properties from ab initio calculations.

Linear elasticity is still used to build the initial configuration, displacing all atoms according to the superposition of the displacement fields created by each dislocation composing the periodic array. The summation on periodic images can be either performed in reciprocal space (Daw 2006) or in direct space after regularization of the conditionally convergent sums (Cai et al. 2003). The crystal orientation used in such elastic calculations should be chosen so as to fix the displacement
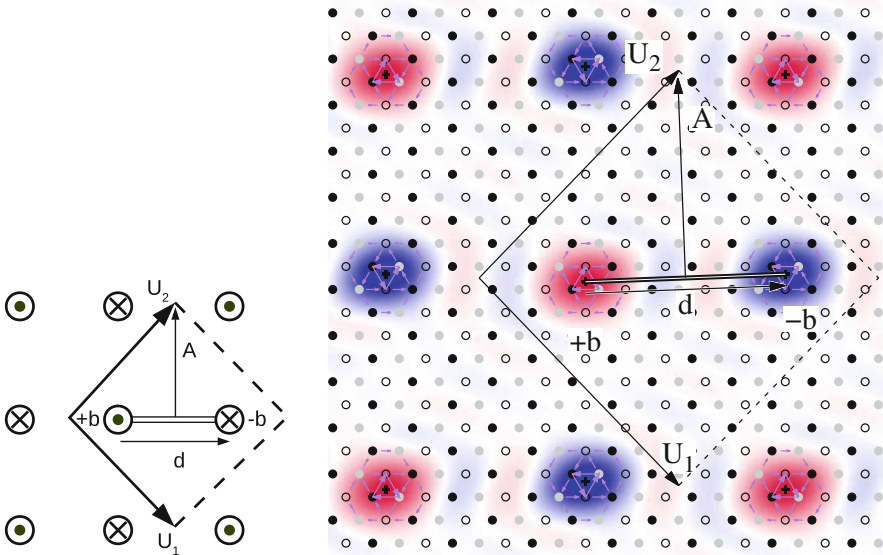
**Fig. 2** Simulation of a dislocation dipole with periodic boundary conditions, using a quadrupolar arrangement. The dipole is defined by its Burgers vector **b**, the dipole vector **d** joining the two dislocation centers, and the cut vector **A**, with the corresponding discontinuity surface indicated by a double black line. $\mathbf{u}_1$ and $\mathbf{u}_2$ are the periodicity vectors of the simulation cell perpendicular to the dislocation line. The example on the right corresponds to the simulation cell used for the modeling of the $1/2 \langle 111 \rangle$ screw dislocation in bcc iron. The dislocation core structures are shown through their differential displacement maps and their density (cf. Fig. 3a for a details)

discontinuity exactly in between the two dislocations composing the dipole, thus preventing the propagation of this discontinuity to infinity. The cut vector **A** defining this discontinuity (Fig. 2) is therefore given by $\mathbf{A} = \mathbf{l} \times \mathbf{d}$, where **l** is the line vector of the dislocations and **d** the vector joining the centers of the $+\mathbf{b}$ dislocation to the $-\mathbf{b}$ one. If the scalar product **A.b** is non-null, i.e., if the dislocation dipole has an edge component and the displacement discontinuity does not coincide with the dislocation glide plane, it is also necessary to insert atoms into or delete them from the original lattice at the discontinuity location, thus following the Volterra operation.

A homogeneous strain needs also to be applied to accommodate the plastic strain created by the dipole (Daw 2006; Cai et al. 2003) and ensure that the average stress in the simulation cell is null. This can be easily demonstrated by considering the variation of the elastic energy when a homogeneous strain $\varepsilon_{ij}$ is applied to a simulation cell containing a dislocation dipole defined by its Burgers vector **b** and its cut vector **A**

$$\Delta E(\varepsilon) = \frac{1}{2} S\, C_{ijkl}\, \varepsilon_{ij}\, \varepsilon_{kl} + C_{ijkl}\, b_i\, A_j\, \varepsilon_{kl},$$

where energies are defined per dislocation unit length and have been thus normalized by the height of the simulation cell in the direction of the dislocation line. $S$ is the area of the simulation cell perpendicular to this direction and $C_{ijkl}$ are the elastic constants. The average stress existing in the simulation cell is then given by

$$\sigma_{ij} = \frac{1}{S} \frac{\partial \Delta E}{\partial \varepsilon_{ij}} = C_{ijkl} \left( \varepsilon_{kl} - \varepsilon_{kl}^0 \right), \tag{2}$$

with the plastic strain defined by

$$\varepsilon_{kl}^0 = -\frac{b_i A_j + b_j A_i}{2S}. \tag{3}$$

One thus sees that the stress given by Eq. 2 is null when the applied strain $\varepsilon_{ij}$ is equal to the plastic strain $\varepsilon_{ij}^0$. When this applied strain is different, a Peach-Koehler force acting on the dislocations may exist. This allows studying properties of the dislocation core under an applied stress, to determine its Peierls stress, for instance. Finally, when a stress variation is observed in ab initio calculations, Eq. 2 allows to deduce the plastic strain from this stress, and thus the dislocations' relative positions via the cut vector **A** (Eq. 3). For instance, the trajectories of the screw dislocations gliding between two neighboring Peierls valleys have been determined, thanks to this method in HCP Zr (Chaari et al. 2014) and in BCC transition metals (Dezerald et al. 2016).

With these periodic boundary conditions, all the excess energy contained in the simulation cell is due to the dislocations. As it will be shown in the last section, elasticity theory can be used to isolate the contribution of a single dislocation. These periodic boundary conditions offer thus a convenient way to extract dislocation energy from ab initio calculations. But the dislocation core structure, and hence the associated excess energy, can be perturbed by the presence of the periodic images. In practice, one will therefore need to check how sensitive are the obtained dislocation properties with the size of the simulation cell.

## 4  Dislocation Core Structures

Different representations can be used to image and analyze the relaxed dislocation core structure obtained by atomic simulations. This allows, for instance, highlighting a spreading of or a dissociation of the dislocation.

### 4.1  Differential Displacement Maps

Differential displacement maps were introduced by Vitek et al. (1970). Two examples are shown in Fig. 3 for a screw dislocation in a body-centered cubic (bcc) crystal and a hexagonal close-packed (hcp) crystal. In these maps, the crystal

is projected in the plane perpendicular to the dislocation line, using for atomic columns the positions in the perfect crystal. The differential displacement caused by the dislocation is calculated by considering the difference between the vector connecting two neighbor atoms in the relaxed dislocated crystal and the same connecting vector in the perfect crystal. One then plots the projection of this differential displacement along the direction of the Burgers vector with an arrow pointing from one atomic column to the other, centered in the middle of the two columns and with an amplitude proportional to the differential displacement. As the arrows are proportional to the displacement difference, they are a representation of the discrete derivative of the displacement field, i.e., of the strain created by the dislocation.

The differential displacement map of the $1/2 \langle 111 \rangle$ screw dislocation in bcc Fe shown in Fig. 3a highlights the compactness and the threefold symmetry of the core. Arrows have been normalized so that an arrow linking the centers of two atomic columns corresponds to a differential displacement of $b/3$. One can thus draw Burgers circuits on this map and obtain the norm of the Burgers vector of the enclosed dislocation by summing arrows. The only non-null Burgers vector is obtained for circuits containing the dislocation center indicated by a cross, in particular for the triangle connecting the three central [111] atomic rows, with a norm equal to $b$. The dislocation is thus well localized.



(a) bcc Fe: $\mathbf{b} = 1/2 \langle 1\,1\,1 \rangle$           (b) hcp Zr: $\mathbf{b} = 1/3 \langle 1\,\overline{2}\,1\,0 \rangle$
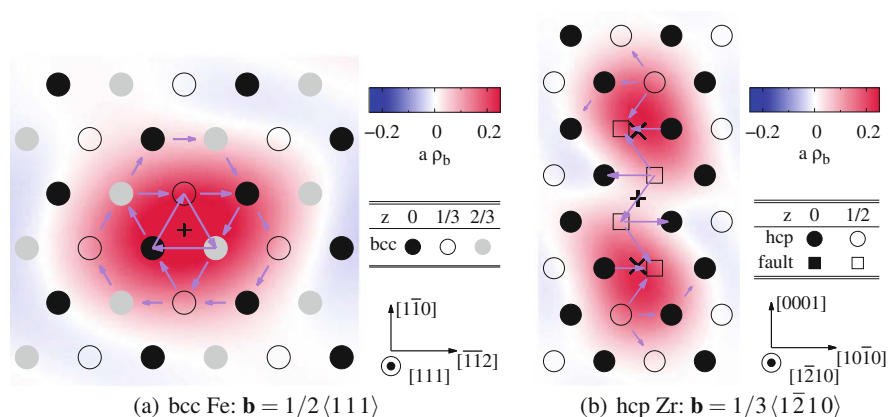
**Fig. 3** Core structure of a $\mathbf{b}$ screw dislocation (**a**) in bcc iron (Dezerald et al. 2016) and (**b**) in hcp zirconium (Clouet et al. 2015). In these projections perpendicular to the dislocation line, atoms are sketched by symbols with a colour depending on their (**a**) (111) and (**b**) ($1\overline{2}10$) plane in the original perfect crystal. In (**b**), different symbols are used for atoms depending on their neighbourhood in the dislocated crystal, i.e. close to the perfect hcp crystal (circles) or to the unrelaxed prismatic stacking fault (squares). The arrows between atomic columns are proportional to the differential displacement created by the dislocation in the direction of the Burgers vector. The colour map show the dislocation density $\rho_b$ normalized by the lattice parameter (Nye tensor). The center of the dislocation is indicated by a + cross. The × crosses in (**b**) correspond to the positions of the partial dislocations deduced from the disregistry in Fig. 4

The picture is quite different for the $1/3\langle 1\bar{2}10\rangle$ screw dislocation in hcp Zr shown in Fig. 3b. The differential displacement map shows a non-isotropic distribution with a spreading of the dislocation core in the $(10\bar{1}0)$ prismatic plane. The normalization here ensures that the maximal arrows correspond to a $b/2$ differential displacement. The presence of a ribbon with arrows having almost the same length therefore corresponds to a $b/2$ prismatic stacking fault which is known to be stable in this transition metal. The differential displacement map thus clearly evidences the dissociation of the screw dislocation in two $1/6\langle 1\bar{2}10\rangle$ partial dislocations separated by a prismatic stacking fault.

## 4.2    Dislocation Density

Another visualization method proposed by Hartley and Mishin (2005) consists of extracting the Nye tensor from the relaxed atomic structure, thus giving a measure of the dislocation density. The component $\alpha_{jk}$ of the Nye tensor corresponds to the density of dislocations with a line direction along $\mathbf{e}_k$ and a Burgers vector along $\mathbf{e}_j$. If $A$ is a surface element of normal $\mathbf{n}$, the dislocation content of line defects along $\mathbf{n}$ intersecting $A$ is given by the surface integral

$$\mathbf{b} = \int_A \alpha \cdot \mathbf{n}\, dS.$$

We only give here the salient points of the method to extract the Nye tensor from atomic simulations, and the reader is referred to the original publication for the practical implementation.

The first step is to define the elastic distortion, i.e., the gradient of the elastic displacement, at each atomic position. Note that this differs from the gradient of the total displacement. One cannot simply compare the atomic positions after and before the introduction of the dislocation to obtain this elastic distortion, but one needs to find for each position the closest undistorted environment corresponding to a zero-stress state. This is performed by comparing, for each atom, the positions of its nearest neighbors in the dislocated relaxed crystal with the ones in a perfect crystal. Knowing the two sets of neighbor positions, each bond in the dislocated crystal, defined by its vector $\mathbf{P}^{(\gamma)}$, is identified with the corresponding $\mathbf{Q}^{(\beta)}$ bond in the perfect crystal, which is the perfect bond leading to the smallest angle $\Phi^{(\gamma\beta)}$ between the vectors $\mathbf{P}^{(\gamma)}$ and $\mathbf{Q}^{(\beta)}$. Only the bonds which are not too much distorted and for which the angle $\Phi^{(\gamma\beta)}$ is smaller than a chosen threshold are kept. The elastic distortion $F^{\mathrm{e}}$ is then locally defined through the relation $P_i^{(\gamma)} = F_{ij}^{\mathrm{e}} Q_j^{(\beta)}$. This cannot be satisfied for each set of associated bond $(\gamma\,\beta)$ as the system of equations is overdetermined and the matrix $F^{\mathrm{e}}$ is obtained by the pseudo-inverse method, i.e., a least-square fitting. The Nye tensor $\alpha$ is then defined through the curl of the inverse transpose of the distortion, $\alpha = -\nabla \times (^+F^{\mathrm{e}})^{-1}$, using finite differences between neighbor atoms for derivation. This defines the Nye tensor on a set of discrete points,

generally atomic positions, which can be then interpolated with cubic splines or Fourier series, or smeared with Gaussian-like spreading functions.

The dislocation density obtained for the $1/2 \langle 111 \rangle$ screw dislocation in bcc Fe (Fig. 3a) illustrates the compactness of the core: the distribution has only one peak. On the other hand, the dislocation distribution for the $1/3 \langle 1\bar{2}10 \rangle$ screw dislocation in hcp Zr (Fig. 3b) shows two well-separated peaks which correspond to the two partial dislocations. To obtain the Nye tensor in this latter case, the neighborhood of each atom in the dislocated crystal is compared not only to the two different neighborhoods existing in the perfect hcp crystal but also to the ones of the unrelaxed prismatic stacking fault, to identify the closer reference from which the elastic distortion is calculated.

## 4.3    Disregistry

The extraction of the disregistry offers another way to characterize the dislocation core structure, particularly convenient when the core is planar. The disregistry is the difference of displacement induced by the dislocation between the plane just above and the one just below the dislocation glide plane. It is thus obtained from the relaxed configuration through

$$\mathbf{D}(x) = \mathbf{u}^{+}(x) - \mathbf{u}^{-}(x),$$

where $\mathbf{u}^{+}(x)$ and $\mathbf{u}^{-}(x)$ are the displacements of the atoms belonging, respectively, to the upper and lower planes and located at the position $x$ in the direction perpendicular to the dislocation line. This disregistry varies from $\mathbf{0}$ for $x \to -\infty$ to $\mathbf{b}$ for $x \to \infty$, thus corresponding to the dislocation glide plane being locally sheared by one Burgers vector $\mathbf{b}$. The dislocation center $x_{\mathrm{D}}$ is defined by $\mathbf{D}(x_{\mathrm{D}}) = \mathbf{b}/2$. The disregistry derivative, $\boldsymbol{\rho}(x) = \partial \mathbf{D}(x)/\partial x$, corresponds to the dislocation density in the glide plane. If the cut plane used to introduce the dislocation in the simulation cell does not correspond to its glide plane, it is necessary to define the atomic displacement in the $\mathbf{0}$ to $\mathbf{b}$ interval. This can be done as the Burgers vector $\mathbf{b}$ of a perfect dislocation is a periodicity vector of the lattice and adding a displacement $n\mathbf{b}$ ($n \in \mathbb{Z}$) to an atom does not change the configuration.

Peierls and Nabarro built a model that leads to a simple analytical expression of the disregistry (Lu 2005). According to this model, the disregistry is given by

$$D(x) = \frac{b}{\pi} \left[ \arctan \left( \frac{x - x_{\mathrm{D}}}{\zeta} \right) + \frac{\pi}{2} \right],$$

where $x_{\mathrm{D}}$ is the dislocation position and $\zeta$ its spreading in the glide plane. For simplicity, we consider scalar quantities by projecting the displacement in the direction of the Burgers vector. Fitting of these two parameters to the data extracted from the atomistic simulations allows thus defining the dislocation position and characterizing the spreading of its core.

For dissociated dislocations, the disregistry is the sum of the contributions of the two partial dislocations, i.e.,, assuming that each partial dislocation has the same Burgers vector $\mathbf{b}/2$ and the same spreading $\zeta$:

$$D(x) = \frac{b}{2\pi} \left[ \arctan\left( \frac{x - x_D - d/2}{\zeta} \right) + \arctan\left( \frac{x - x_D + d/2}{\zeta} \right) + \pi \right],$$

where $d$ is the dissociation distance. As shown in Fig. 4 for the $1/3 \langle 1\bar{2}10 \rangle$ screw dislocation in hcp Zr, such an analytical expression generally perfectly describes the disregistry extracted from the atomic simulations. One can also notice that the positions in the glide plane of the partial dislocations deduced from the disregistry agree which what can be inferred from the differential displacement and the Nye tensor maps (Fig. 3b). Some consequences of the periodic boundary conditions used to model this dislocation are visible on these disregistry plots. The dislocation density slightly depends, through the dissociation distance $d$ and the partial spreading $\zeta$, on the simulation cell, not only its size but also its shape. One also sees that the density of the periodic dislocation arrays (solid line in Fig. 4), obtained by summation of the contributions of the periodic images in the glide plane, slightly differs from the one of the isolated dislocation (dashed line in
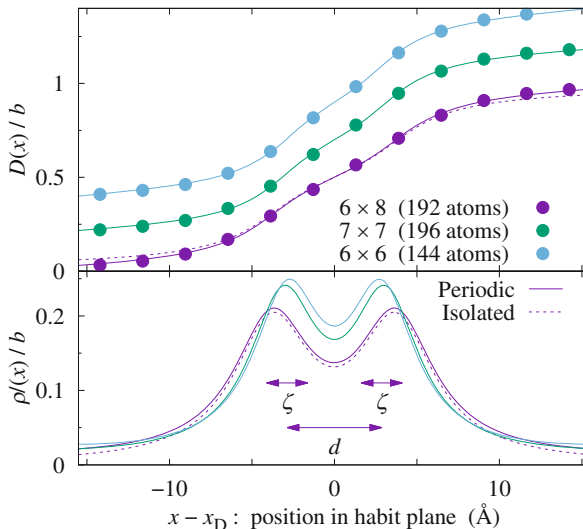


**Fig. 4** Disregistry $D(x)$ created by a $1/3 \langle 1\bar{2}10 \rangle$ screw dislocation in its $(10\bar{1}0)$ prismatic glide plane in hcp Zr, and corresponding dislocation density $\rho(x) = \partial D(x)/\partial x$. Symbols correspond to ab initio calculations and lines to the fit of the Peierls-Nabarro model, considering periodicity or not (straight and dashed lines respectively). Results are shown for different $n \times m$ periodic arrangements corresponding to the periodicity vectors $\mathbf{u}_1 = n/2 \,[10\bar{1}0]$ and $\mathbf{u}_2 = m \,[0001]$ (see Clouet 2012 for details). For clarity, disregistries $D(x)$ have been shifted by 0.2 between different data sets. The obtained dissociation distance $d$ and spreading $\zeta$ of the partial dislocations are indicated for the $6 \times 8$ periodic array whose core structure is shown in Fig. 3b

Fig. 4), especially in the distribution tail. Flexible boundary conditions, as discussed in Sect. 3.2, have been developed to solve such limitations of periodic boundary conditions.

## 5    Dislocation Energy

Ab initio calculations give access to the dislocation core energy and its variations. This core energy is the part of the dislocation excess energy which arises from the strong perturbation of the atomic interactions in the immediate vicinity of the dislocation line and which cannot be described by linear elasticity. Contrary to the dislocation elastic energy, this is an intrinsic property which only depends on the dislocation and not on the surrounding environment. When several configurations exist for the same dislocation, this core energy controls their relative stability. Its variations with the position of the dislocation in the crystal lattice is at the origin of the lattice friction opposing dislocation glide.

### 5.1    Core Energy

Among the different boundary conditions introduced in Sect. 3 to model a dislocation at an atomic scale, only periodic boundary conditions allow for an unambiguous determination of the dislocation core energy with ab initio calculations. This is a consequence of the energy formulation inherent to ab initio calculations. Because of the non-locality of the electronic energy, which contains a contribution which needs to be evaluated in reciprocal space, one cannot easily partition the excess energy of the simulation cell between the dislocation and the external boundary contributions when a defective boundary has been introduced like in cluster approaches using either fixed (Sect. 3.1) or flexible boundaries (Sect. 3.2). Ab initio methods to project the energy on atoms have been proposed: they theoretically allow for such a partition but the application to the calculation of a dislocation core energy still remains to be done. Even if the absolute value of the core energy appears difficult to determine with cluster approaches, methods to estimate its variation are nevertheless possible. One can, for instance, calculate the difference of core energy between two configurations of the same dislocations by simply considering the difference of ab initio total energies. But such an approach assumes that the contribution of the external boundary will cancel in the difference, an assumption which may be hard to validate. Variation of the dislocation energy with its position in the crystal lattice can also be estimated by considering the work of the atomic forces during the motion (Swinburne and Kermode 2017).

On the other hand, with periodic boundary conditions, all the excess energy arises from the dislocations. This excess energy $\Delta E$ is defined as the energy difference per unit of height between the supercell with and without the dislocation dipole. If atoms have been removed or inserted during the creation of the dipole, the energy of the perfect supercell needs to be normalized by the correct number of atoms. It is given

by the sum of the core energy $E^{\text{core}}$ of the two dislocations, of the elastic energy $E^{\text{elas}}_{\text{dipole}}$ of the dipole contained in the supercell and of its elastic interaction with its periodic images:

$$\Delta E = 2\, E^{\text{core}} + E^{\text{elas}}_{\text{dipole}} + \frac{1}{2} \sum_{n,m} E^{\text{elas}}_{\text{inter}} (n\, \mathbf{u}_1 + m\, \mathbf{u}_2). \tag{4}$$

The factor $1/2$ appears in front of this last contribution as only one half of the interaction is attributed to each interacting dipole. When partitioning the excess energy into a core and an elastic contribution, it is necessary to introduce a cutoff distance to isolate the dislocation cores. Close to the dislocation lines, strains are much too high to be described by linear elasticity. As a consequence, elastic fields diverge at the origin, and one needs to exclude the core region from the elastic description. The elastic contribution to the excess energy is thus obtained by integrating the elastic energy density on the whole supercell except two cylinders of radius $r_{\text{c}}$ which isolate this elastic divergence. The core energy corresponds to the excess energy contained in these cylinders. The total excess energy $\Delta E$ does not depend on the choice for this core radius, but the partition between a core and an elastic contribution depends on $r_{\text{c}}$.

The elastic energy of the dipole and its interaction with its periodic images can be computed by considering the Volterra elastic field created by the dislocations. This calculation can be performed either in reciprocal space (Daw 2006) or in direct space using classical results of dislocation elastic theory (Bacon et al. 1980). In this last case, one uses the decomposition of Eq. (4), with the contribution of the dipole contained in the supercell and its interaction with the periodic images calculated separately. The dipole elastic energy is obtained by the volume integral:

$$E^{\text{elas}}_{\text{dipole}} = \frac{1}{2} \iiint_V \left( \sigma^{(1)}_{ij} + \sigma^{(2)}_{ij} \right) \left( \varepsilon^{(1)}_{ij} + \varepsilon^{(2)}_{ij} \right) dV,$$

where $\sigma^{(n)}$ and $\varepsilon^{(n)}$ are the stress and strain created by the dislocation $n$. This is transformed into a surface integral, thanks to Gauss' theorem:

$$E^{\text{elas}}_{\text{dipole}} = \frac{1}{2} \iint_S \left( \sigma^{(1)}_{ij} + \sigma^{(2)}_{ij} \right) \left( u^{(1)}_i + u^{(2)}_i \right) dS_j,$$

with $\mathbf{u}^{(n)}$ the displacement field associated with dislocation $n$. The integration surface is composed of the two cylinders $S^{(1)}_{\text{c}}$ and $S^{(2)}_{\text{c}}$ of radii $r_{\text{c}}$ removing the elastic divergence at the dislocation cores, and of the two surfaces $S_{0-}$ and $S_{0+}$ removing the displacement discontinuity along the dislocation cut (Fig. 5). The integration on both core cylinders leads to the same contribution

$$E^{\text{elas}}_{\text{c}}(\phi) = \frac{1}{2} \iint_{S^{(1)}_{\text{c}}} \sigma^{(1)}_{ij} u^{(1)}_i dS_j = \frac{1}{2} \iint_{S^{(2)}_{\text{c}}} \sigma^{(2)}_{ij} u^{(2)}_i dS_j. \tag{5}$$

**Fig. 5** Definition of the contour surface used to calculate the elastic energy of a dislocation dipole
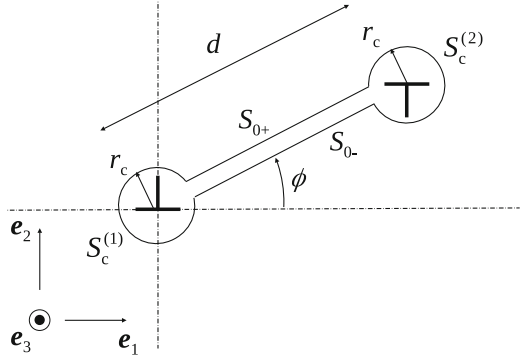


This contributions of the core tractions to the elastic energy (Clouet 2009) should not be forgotten as it ensures that the elastic energy is a state variable compatible with the work of the Peach-Koehler forces. Besides, in ab initio calculations where the distance $d$ between the two dipole dislocations is small, this can lead to a non-negligible contribution compared to the one associated with the integral along the cut surface, even for a screw orientation. The elastic energy of the dislocation dipole is then

$$E_{\text{dipole}}^{\text{elas}} = 2E_{\text{c}}^{\text{elas}}(\phi) + b_i K_{ij}^0 b_j \ln\left(\frac{d}{r_{\text{c}}}\right), \tag{6}$$

where the tensor $\mathbf{K}^0$ defined by Stroh (1958, 1962) only depends on the elastic constants. The total elastic energy is finally obtained by adding the interaction of the dipole with its periodic images. But, one should realize that the summation on periodic images appearing in Eq. (4) is only conditionally convergent: it can be regularized with the method of Cai et al. (2003).

As shown in Fig. 6 for the $1/2 \langle 111 \rangle$ screw dislocation in bcc iron, once this elastic energy is subtracted from the dislocation excess energy given by ab initio calculations, one obtains a constant core energy which does not depend on the size of the supercell. Some slight variations of the core energy are nevertheless still observed with the type of periodic arrangement used for the atomic simulations. These variations arise because only the Volerra elastic field has been considered in the calculation of the elastic energy. Dislocations also cause a core elastic field, which decays more rapidly than the Volterra elastic field. Because of the small size of the supercell used in ab initio calculations, this core field may also lead to an elastic interaction between the different dislocations composing the periodic array. This contribution to the elastic energy can be computed to improve the convergence of core energies (Clouet et al. 2009). A quadrupolar periodic arrangement minimizes this contribution of the core field. This is why such an arrangement is preferred when periodic boundary conditions are used. The neglect of anharmonic effects in the calculation of the elastic energy can also be a reason for the variation of the core energy with the supercell. Knowing higher-order elastic constants, one can

**Fig. 6** Decomposition of the excess energy $\Delta E$ of a $1/2 \langle 111 \rangle$ screw dislocation dipole in bcc Fe in an elastic contribution $E^{\text{elas}}$ and a core energy $E^{\text{core}}$, using a core radius $r_c = b/2$. Different symbols correspond to different periodic arrangements (see Clouet et al. 2009 for details)
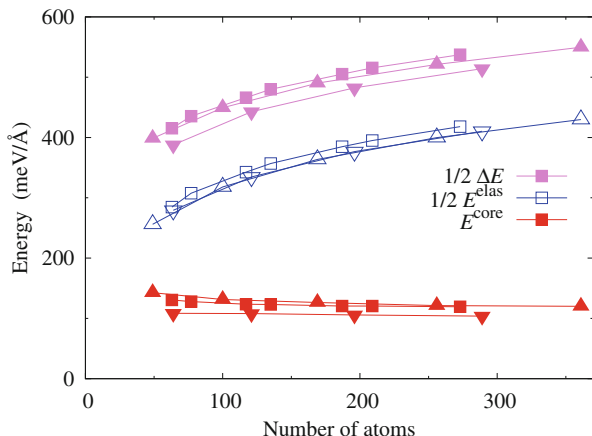


use nonlinear elasticity theory in principle to calculate more precisely this elastic contribution (Teodosiu 1982). But this leads to much cumbersome calculations. In practice, as anharmonicity is important only close to the dislocation core, the consideration of the dislocation core field offers a way to incorporate anharmonic effects while still relying on linear elasticity.

## 5.2   Peierls Energy Barrier

The Peierls energy is the energy barrier opposing dislocation glide. It corresponds to a variation of the dislocation core energy as the elastic energy is not dependent upon the dislocation position in the crystal lattice. It can be calculated by finding the minimum energy path linking two neighboring stable positions of the dislocation using either constrained minimization or nudged elastic band (NEB) calculations (Henkelman et al. 2000).

With periodic boundary conditions, the Peierls energy is directly obtained by considering a path where both dislocations composing the dipole are displaced by one Peierls valley in the same direction. If the two dislocations are moved simultaneously along the path, their separation distance does not vary, and the elastic energy is constant. This ensures that the energy variation given by the constrained minimization or the NEB calculations directly corresponds to the Peierls energy. However, this is possible only if crystal symmetry ensures that the path is symmetrical as the two dislocations are traversing their Peierls barriers in the opposite direction. This is the case, for instance, for the $1/2 \langle 111 \rangle$ screw dislocation in a bcc lattice gliding in a $\{110\}$ plane (Fig. 7a).

If the path is not symmetrical, either because of the lack of crystal symmetries or because of an applied stress, it is not possible anymore to move both dislocations simultaneously in the same direction. One needs either to move them in opposite directions or to keep one dislocation fixed when the second one is moving. As
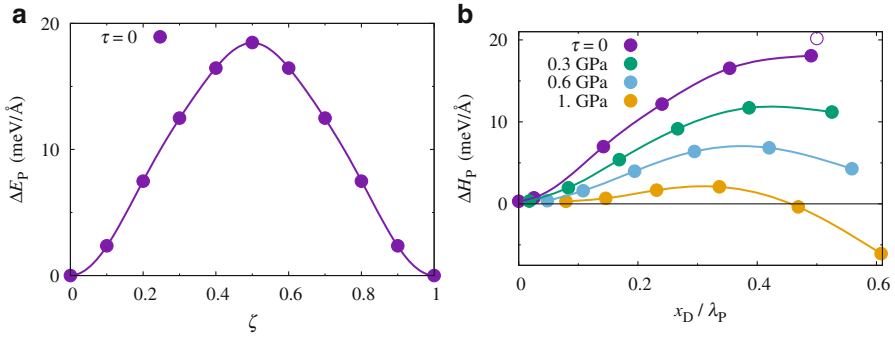
**Fig. 7** Peierls barrier of a $1/2 \langle 111 \rangle$ screw dislocation in bcc Mo. (**a**) The energy variation $\Delta E$ is shown as a function of the reaction coordinate $\zeta$, and (**b**) the enthalpy variation $\Delta H$ as a function of the dislocation position $x_{\mathrm{D}}$ normalized by the distance $\lambda_{\mathrm{P}}$ between two Peierls valleys (see Dezerald et al. 2014, 2016 for details). For the Peierls barriers under stress (**b**), only one half of the barrier has been computed, with one dislocation of the dipole being displaced while the second one remains fixed. The open symbol is the enthalpy variation in the middle of the pathway ($x_{\mathrm{D}}/\lambda_{\mathrm{P}} = 1/2$) before correcting for the variation of the elastic interaction energy for the $\tau = 0$ calculation

a consequence, the separation distance, and thus the elastic interaction energy, is varying along the path. One can calculate this variation of the elastic energy and subtract it from the excess energy in order to obtain the Peierls energy. To be able to perform this elastic calculation, one needs first to determine the exact dislocation position $x_{\mathrm{D}}$ for each reaction coordinate $\zeta$ along the path. This can be done using the dislocation disregistry (cf. Sect. 4.3) if the motion is planar or by fitting the atomic displacements with the Volterra elastic solution. As the stress is directly linked to the applied strain and the dislocation positions (Eqs. 2 and 3), one can also use the stress variation observed along the dislocation path to determine the dislocations position. The example of Fig. 7b shows that, with this correction for the variation of the elastic energy, the same Peierls energy is obtained under zero applied stress when one dislocation is fixed or when both dislocations are moved (Fig. 7a).

## 5.3    Peierls Stress

The Peierls stress is the applied resolved shear stress necessary to cancel the Peierls barrier so that the dislocations can glide freely without the need of thermal activation, i.e., the stress necessary to move the dislocation at 0 K. For an applied stress $\tau$, the Peierls barrier is given by the enthalpy variation

$$\Delta H_{\mathrm{P}}(x_{\mathrm{D}}, \tau) = \Delta E_{\mathrm{P}}(x_{\mathrm{D}}) - \tau b x_{\mathrm{D}},$$

which corresponds to the Peierls energy barrier plus the work of the applied stress when the dislocation has glided a distance $x_{\mathrm{D}}$. The Peierls stress is thus

the maximum applied stress $\tau$ for which the function $\Delta H_P(x_D, \tau)$ goes through a maximum in the range $0 \leq x_D \leq \lambda_P$. If one assumes that the energy barrier $\Delta E_P(x_D)$ does not depend on the applied stress $\tau$, it is given by

$$\tau_P = \frac{1}{b}\text{Max}\left(\frac{\partial \Delta E_P}{\partial x_D}\right). \tag{7}$$

The Peierls stress can thus be theoretically obtained from the calculation of the Peierls energy barrier under zero applied stress. But, the evaluation of the derivative in Eq. 7 requires to know the variation of the energy as a function of the dislocation position and not only of the reaction coordinate. In practice, the obtained value for $\tau_P$ will sensitively vary with the method chosen to estimate the dislocation position along the path.

One can also directly calculate with ab initio calculations the Peierls barrier under an applied stress so as to estimate at which stress the barrier cancels (Fig. 7). In such calculations, one does not really apply a stress but a strain corresponding to the target stress (Eq. 2). With periodic boundary conditions, as the distance between the two dislocation is varying, the applied stress is also varying along the path. Equations (2) and (3) show that the stress variation is directly proportional to the dislocation displacement and to the inverse of the surface $S$ of the simulation cell perpendicular to the dislocation line. If only one dislocation is moving along the path, this stress variation therefore does not exceed

$$\delta\tau = \mu\frac{b\lambda_P}{S},$$

where $\mu$ is the shear modulus in the dislocation glide plane.

If one is only interested in the calculation of the Peierls stress and not in the variation of the Peierls barrier with the applied stress, one can simply perform static relaxation of a dislocation under an applied stress to see at which applied stress the dislocation glides by at least one Peierls valley. With periodic boundary conditions one still needs to take into account the variation of the elastic interaction and of the applied stress when the dislocation is moving to interpret the results. On the other hand, no such artifact exists with a cluster approach using flexible boundary conditions which truly models a single isolated dislocation under an applied stress. Straining homogeneously the simulation cell to obtain the targeted applied stress, the Peierls stress is defined as the stress for which the dislocation cannot be stabilized anymore and escapes from the cluster. If one is only interested in the evolution of the dislocation core structure under an applied stress and on the determination of the Peierls stress, this cluster approach therefore appears as the method of choice. Nevertheless, whatever the boundary conditions, determination of the Peierls stress by such an instability condition of the dislocation core under an applied stress necessitates a strict threshold criterion on the atomic forces to obtain a meaningful value.

# 6    Conclusions

Dislocation core properties can now be routinely studied with ab initio calculations, thanks to the different methodological developments summarized in this chapter. This usually necessitates a coupling between atomistic model and elasticity theory, for which different already available tools can be used: see, for instance, D. R. Trinkle website (http://dtrinkle.matse.illinois.edu) for an implementation of the lattice Greens functions or the Babel package (http://emmanuel.clouet.free.fr/Programs/Babel) for handling dislocations in atomistic simulation cells and elastic energy calculations. Useful information on the dislocation core structure are thus obtained. Such calculations can, for instance, characterize possible dissociation or spreading of the core, or evidence the existence of several stable configurations for the same dislocation. One gets access to the different energy barriers opposing the dislocation motion and to their variation with the applied stress. It is also possible to study how these core properties are altered by the interaction with solute atoms.

Because of the limited size that can be handled by ab initio calculations, such studies are usually limited to the study of straight dislocation, and only few ab initio calculations have considered until now the presence of kinks on the dislocation lines. Upscaling modeling approaches, relying, for instance, to the line tension approximation to describe kink nucleation, are therefore needed to go from these fundamental core properties determined at $0\,\mathrm{K}$ with ab initio calculations to dislocation mobility laws at finite temperature. Larger atomistic simulations are also possible using empirical potentials to describe atomic interactions. These simulations allow studying more complex situations and simulating different dislocation mechanisms, like glide, cross-slip and interaction with other elements of the microstructure, without assuming a priori the elementary mechanism. In such a context, ab initio calculations are useful to validate and also help the development of empirical potentials which correctly reproduce dislocation fundamental properties.

# References

Bacon DJ, Barnett DM, Scattergood RO (1980) Anisotropic continuum theory of lattice defects. Prog Mater Sci 23:51–262. https://doi.org/10.1016/0079-6425(80)90007-9

Cai W, Bulatov VV, Chang J, Li J, Yip S (2003) Periodic image effects in dislocation modelling. Philos Mag 83:539–567. https://doi.org/10.1080/0141861021000051109

Chaari N, Clouet E, Rodney D (2014) First-principles study of secondary slip in zirconium. Phys Rev Lett 112:075504. https://doi.org/10.1103/PhysRevLett.112.075504

Chen Q, Liu XY, Biner S (2008) Solute and dislocation junction interactions. Acta Mater 56:2937–2947. https://doi.org/10.1016/j.actamat.2008.02.026

Choly N, Lu G, Weinan E, Kaxiras E (2005) Multiscale simulations in simple metals: a density-functional-based methodology. Phys Rev B 71:094101. https://doi.org/10.1103/physrevb.71.094101

Clouet E (2009) Elastic energy of a straight dislocation and contribution from core tractions. Philos Mag 89:1565–1584. https://doi.org/10.1080/14786430902976794

Clouet E (2012) Screw dislocation in zirconium: an ab initio study. Phys Rev B 86:144104. https://doi.org/10.1103/PhysRevB.86.144104

Clouet E, Ventelon L, Willaime F (2009) Dislocation core energies and core fields from first principles. Phys Rev Lett 102:055502. https://doi.org/10.1103/PhysRevLett.102.055502

Clouet E, Caillard D, Chaari N, Onimus F, Rodney D (2015) Dislocation locking versus easy glide in titanium and zirconium. Nat Mater 14:931–936. https://doi.org/10.1038/nmat4340

Daw MS (2006) Elasticity effects in electronic structure calculations with periodic boundary conditions. Comput Mater Sci 38:293–297. https://doi.org/10.1016/j.commatsci.2006.02.009

Dezerald L, Ventelon L, Clouet E, Denoual C, Rodney D, Willaime F (2014) Ab initio modeling of the two-dimensional energy landscape of screw dislocations in BCC transition metals. Phys Rev B 89:024104. https://doi.org/10.1103/PhysRevB.89.024104

Dezerald L, Rodney D, Clouet E, Ventelon L, Willaime F (2016) Plastic anisotropy and dislocation trajectory in BCC metals. Nat Commun 7:11695. https://doi.org/10.1038/ncomms11695

Eshelby JD, Read WT, Shockley W (1953) Anisotropic elasticity with applications to dislocation theory. Acta Metall 1:251–259. https://doi.org/10.1016/0001-6160(53)90099-6

Hartley CS, Mishin Y (2005) Characterization and visualization of the lattice misfit associated with dislocation cores. Acta Mater 53:1313–1321. https://doi.org/10.1016/j.actamat.2004.11.027

Henkelman G, Jóhannesson G, Jónsson H (2000) Methods for finding saddle points and minimum energy paths: theoretical methods in condensed phase chemistry. In: Schwartz SD (ed) Progress in theoretical chemistry and physics, vol 5, chap 10. Springer, Dordrecht (NL), pp 269–302. https://doi.org/10.1007/0-306-46949-9_10

Hirth JP, Lothe J (1982) Theory of dislocations, 2nd edn. Wiley, New York

Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864–B871. https://doi.org/10.1103/PhysRev.136.B864

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlations effects. Phys Rev 140:A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Liu Y, Lu G, Chen Z, Kioussis N (2007) An improved QM/MM approach for metals. Modelling Simul Mater Sci Eng 15:275–284. https://doi.org/10.1088/0965-0393/15/3/006

Lu G (2005) The Peierls-Nabarro model of dislocations: a venerable theory and its current development. In: Yip S (ed) Handbook of materials modeling, Springer, Dordrecht (NL), pp 793–811

Rao S, Hernandez C, Simmons JP, Parthasarathy TA, Woodward C (1998) Green's function boundary conditions in two-dimensional and three-dimensional atomistic simulations of dislocations. Philos Mag A 77:231–256. https://doi.org/10.1080/01418619808214240

Rodney D, Ventelon L, Clouet E, Pizzagalli L, Willaime F (2017) Ab initio modeling of dislocation core properties in metals and semiconductors. Acta Mater 124:633–659. https://doi.org/10.1016/j.actamat.2016.09.049

Segall DE, Strachan A, Goddard WA, Ismail-Beigi S, Arias TA (2003) Ab initio and finite-temperature molecular dynamics studies of lattice resistance in tantalum. Phys Rev B 68:014104. https://doi.org/10.1103/PhysRevB.68.014104

Sinclair JE, Gehlen PC, Hoagland RG, Hirth JP (1978) Flexible boundary conditions and nonlinear geometric effects in atomic dislocation modeling. J Appl Phys 49:3890–3897. https://doi.org/10.1063/1.325395

Stroh AN (1958) Dislocations and cracks in anisotropic elasticity. Philos Mag 3:625–646. https://doi.org/10.1080/14786435808565804

Stroh AN (1962) Steady state problems in anisotropic elasticity. J Math Phys 41:77

Swinburne TD, Kermode JR (2017) Computing energy barriers for rare events from hybrid quantum/classical simulations through the virtual work principle. Phys Rev B 96:144102. https://doi.org/10.1103/PhysRevB.96.144102

Tan AMZ, Trinkle DR (2016) Computation of the lattice Green function for a dislocation. Phys Rev E 94:023308. https://doi.org/10.1103/PhysRevE.94.023308

Teodosiu C (1982) Elastic models of crystal defects. Springer, Berlin-Heidelberg

Vitek V, Perrin RC, Bowen DK (1970) The core structure of $1/2\langle 111\rangle$ screw dislocations in b.c.c. crystals. Philos Mag 21:1049–1073. https://doi.org/10.1080/14786437008238490

Woodward C (2005) First-principles simulations of dislocation cores. Mater Sci Eng A 400–401:59–67. https://doi.org/10.1016/j.msea.2005.03.039

Yasi JA, Hector LG, Trinkle DR (2012) Prediction of thermal cross-slip stress in magnesium alloys from a geometric interaction model. Acta Mater 60:2350–2358. https://doi.org/10.1016/j.actamat.2012.01.004

# Modeling the Thermally Activated Mobility of Dislocations at the Atomic Scale

# 71

Laurent Proville and David Rodney

## Contents

**Abstract**

We review in this chapter how to model the mobility of isolated dislocations at the atomic scale when glide requires to overcome energy barriers and is thermally activated, as is typically the case in body-centered cubic metals. We first recall the boundary and loading conditions used to model an isolated dislocation. We then detail a static approach based on the Transition State Theory parameterized on atomistic calculations to predict dislocation mobility. Finally, we address the low-temperature regime and explain how to include quantum corrections to the dislocation mobility law.

L. Proville
DEN-Service de Recherche de Metallurgie Physique, CEA, Universite Paris-Saclay, F-91191, Gif sur Yvette, France
e-mail: laurent.proville@cea.fr

D. Rodney (✉)
Institut Lumière Matière, Université Lyon 1, Villeurbanne CEDEX, France
e-mail: david.rodney@univ-lyon1.fr

# 1    Introduction

Dislocations are set in motion under the action of an applied resolved shear stress. In absence of obstacles, that is, if the crystal is locally free of crystalline defects such as other dislocations, precipitates, or grain boundaries, the dislocation may reach a steady-state velocity if the applied stress is constant. In this stationary state, the work of the applied stress, which drives dislocation motion, is balanced by the energy lost to the surrounding crystal as heat (Leibfried 1950; Nadgornyi 1988). The mobility law, i.e., the relation between the applied resolved shear stress and the steady-state velocity, is one of the most fundamental ingredients of all dislocation dynamics theory (Suzuki et al. 1991; Cai and Bulatov 2004; Bulatov and Cai 2006). This relation has been measured experimentally (Johnston and Gilman 1959; Caillard 2010; Nosenko et al. 2011), but there is also an intense research activity to predict the mobility law from elementary dislocation properties (Koizumi et al. 2002). Such models must necessarily include the atomic structure of the dislocation core because the velocity reached by a dislocation at a given stress is intimately related to how its core interacts with the surrounding crystal. In the present chapter, we will discuss how to measure and predict a dislocation mobility law using atomic-scale simulations.

The mobility law concerns the motion of isolated dislocations gliding freely between obstacles. It should not be confused with the long timescale average velocity $v_d$, as estimated from Orowan's law $\dot{\varepsilon}_p = \rho_m b v_d$, where $\dot{\varepsilon}_p$ is the applied strain rate, $\rho_m$ the density of mobile dislocations, and $b$ the Burgers vector. Indeed, a dislocation meets numerous obstacles during its glide in a crystal and undergoes a stop-and-go motion, spending long periods of time immobile in contact with obstacles before unpinning and traveling rapidly to the next obstacle. The average velocity $v_d = \dot{\varepsilon}_p / \rho_m b$, which includes both the pinned and unpinned regimes, can therefore be significantly smaller than the unpinned velocity between obstacles, which is the object of the mobility law. Predicting the average dislocation velocity requires large-scale models, such as discrete dislocation dynamics simulations, to account for the collective interactions between dislocations and between dislocations and the other elements of the microstructure (Bulatov and Cai 2006; Kubin 2013). In contrast, in the present chapter, we consider the kinematic response of an isolated dislocation, which can be studied in simulation cells compatible with the size limit inherent to atomic-scale calculations.

The dislocation mobility has two very distinct regimes, depending on whether the applied stress is below or above the Peierls stress, the intrinsic resistance of the crystal against dislocation motion. If the applied stress is below the Peierls stress, the dislocation cannot glide without the help of thermal fluctuations and jumps between equilibrium positions by a thermally activated process (Caillard and Martin 2003). The resulting dislocation velocity is then thermally activated and depends exponentially on the temperature and the applied stress. As a result, the dislocation kinetics is too slow to be modeled by direct molecular dynamics (MD) simulations, except at high applied stresses close to the Peierls stress or high temperatures (Marian et al. 2004; Chaussidon et al. 2006; Gilbert et al. 2011). It can however

be approximated using the transition state theory (TST) parameterized on atomistic static calculations. This is typically the case of $a/2\langle 111 \rangle$ screw dislocations in body-centered cubic (BCC) metals, like Mo, Ta, or W, which will be discussed in the present chapter.

In the second regime, the applied stress is larger than the Peierls stress. This is typically the case of face-centered cubic (FCC) metals, like Al, Cu, and Ni, where the Peierls stress of the conventional $a/2\langle 110 \rangle$ dislocations is very small, a few tens of MPa at most. In any practical condition of deformation at finite strain rates, the applied stress is then much larger than the Peierls stress. In this stress regime, the dislocation velocity is large enough to be measured directly by MD simulations (Gumbsch and Gao 1999; Rodney and Martin 2000; Bhate et al. 2002; Olmsted et al. 2005; Mordehai et al. 2003; Oren et al. 2017). At stresses slightly greater than the Peierls stress, the dislocation velocity is limited by the rate at which atomic and electronic vibrations can dissipate the heat produced by the work of the applied stress. In MD simulations where only atomic vibrations are accounted for, the velocity increases linearly with the ratio $\tau/T$ where $\tau$ is the applied stress and $T$ the temperature, as expected from theoretical analysis of phonon damping (Leibfried 1950; Alshits 1992). At higher stresses, the dislocation velocity enters a relativistic regime and approaches asymptotically the shear wave speed. Finally, supersonic and transonic velocities may be reached at even higher stresses in conditions akin to shocks. We will not discuss these results further, but rather refer the reader, for example, to the recent work of Oren et al. (2017) and references therein for more details.

In the present chapter, we present how to predict the dislocation velocity using atomistic simulations in the thermally activated regime where the applied stress is below the Peierls stress. Because of the requirements in terms of space scale of these simulations, we will use atomistic simulations based on empirical potentials of the embedded atom method (EAM) type. In Sect. 2, we present how to set up a simulation cell to model an isolated dislocation. In Sects. 3, 4, and 5, we use BCC metals to illustrate how dislocation mobility can be modeled using the transition state theory, both in the high-temperature regime of classical dynamics and at lower temperatures where quantum zero-point energy vibrations can affect the motion of dislocations.

## 2 Construction of the Simulation Cell

There are two main types of boundary conditions adapted to the study of dislocation mobility: periodic or slab conditions. In the former case, a dipole of dislocations of opposite Burgers vectors is introduced in the cell to produce a zero global Burgers vector, which is the condition to apply periodic boundary conditions in all three directions of space. The dislocations of the dipole interact with one another, as well as with their periodic images. The elastic part of these interactions can be corrected, as detailed in Chapter **AB - initio models of dislocations**, yielding well-defined dislocation core energies even in small simulation cells. Periodic boundary
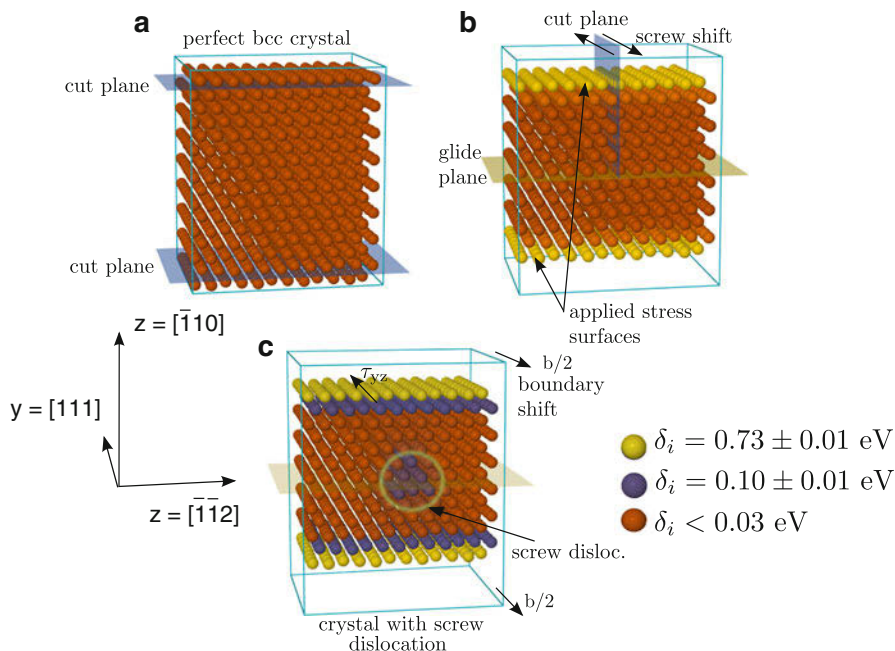
**Fig. 1** Simulation cell construction for a $a/2[111](\bar{1}10)$ screw dislocation in a BCC crystal: (**a**) Perfect crystal simulation cell. (**b**) Simulation cell with free surfaces orthogonal to the Z direction (not relaxed). (**c**) Relaxed simulation cell after introduction of a screw dislocation. Atomic colors are fixed according to the potential energy deviations $\delta_i$ from the perfect crystal cohesive energy (see text)

conditions have been used to study dislocation glide in FCC crystals above the Peierls stress (Mordehai et al. 2003; Oren et al. 2017). One difficulty is that under the action of an applied shear stress, the two dislocations of the dipole move in opposite directions. Their relative distance and mutual interactions vary with time. As a consequence, the velocity of the dislocations vary periodically as they cross the simulation cell (Oren et al. 2017).

The second option is to use a slab geometry with a single dislocation in the cell. An example is shown in Fig. 1 for a screw dislocation in a BCC crystal and in Fig. 2 for an edge dislocation in a FCC crystal. The detailed methodology to construct a slab geometry is described in earlier works (Rodney and Martin 2000; Patinet and Proville 2008; Bacon et al. 2008; Rodney and Proville 2009). The simulation cell is oriented such that the dislocation glide plane is horizontal. In case of a BCC crystal, the crystal orientation is $[\bar{1}\bar{1}2]$ along the $X$ axis, $[111]$ along the $Y$ axis, and $[\bar{1}10]$ along the $Z$ axis. In case of a FCC crystal, we have $[1\bar{1}\bar{2}]$ along the $X$ axis, $[110]$ along the $Y$ axis, and $[1\bar{1}1]$ along the $Z$ axis. The Burgers vector $b = \frac{a}{2}[110]$ in the BCC crystal and $b = \frac{a}{2}[111]$ in the FCC crystal corresponds either to the line direction of the screw dislocation or the glide direction of the edge dislocation. Here, $a$ is the lattice parameter of the perfect lattice.

Starting from a perfect crystal, two free surfaces are created in the direction perpendicular to the glide plane. A dislocation is then introduced by displacing all atoms according to the corresponding isotropic elastic solution (Hirth and Lothe 1982). The anisotropic solution can also be used but is not necessary because the energy in the cell is relaxed after introduction of the dislocation and this relaxation recovers the anisotropy. In order to model an infinite glide plane, periodic boundary conditions are imposed along both the $X$ and $Y$ directions. When a screw dislocation is considered (Fig. 1), an additional shift of half a Burgers vector is introduced in the $Y$ direction (Rodney 2004; Bacon et al. 2008) in order to account for the plastic strain associated with the dislocation and to reconnect the left- and right-$Y$ surfaces. The simulation cell is therefore not rectangular. In the case of an edge dislocation, a half crystal plane orthogonal to the $Y$ direction is added to form a step, as shown in Fig. 2. This step is pushed in the simulation cell when the elastic displacements are applied to form the extra half plane of the edge dislocation. The size of the simulation cell is then increased by half a Burgers vector in the $Y$ direction, again in order to account for the plastic strain of the dislocation (Rodney and Martin 2000; Bacon et al. 2008).

Because of the non-zero Burgers vector in the cell, periodic boundary conditions cannot be applied in the $Z$ direction. Instead, two surfaces are created, forming a slab as shown in Figs. 1 and 2. The simplest solution is to consider free surfaces in this direction, but there are other possibilities, such as fully or semirigid
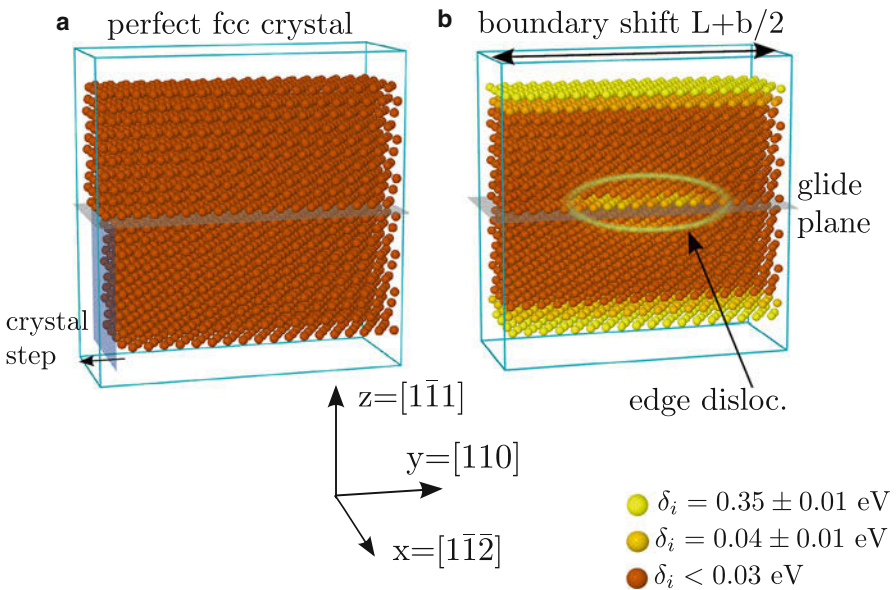


**Fig. 2** Simulation cell construction for a $a/2[110](1\bar{1}1)$ edge dislocation in a FCC crystal: (**a**) Perfect crystal simulation cell with one half plane removed (not relaxed). (**b**) Relaxed simulation cell after the introduction of an edge dislocation. Atomic colors are fixed according to the potential energy deviations from the perfect crystal cohesive energy

boundary conditions, as well as two-dimensional dynamics (Rodney and Martin 2000; Osetsky and Bacon 2003; Bacon et al. 2008). In case of free surfaces, a constant external shear stress $\tau_{xz}$ for the edge dislocation and $\tau_{yz}$ for the screw dislocation may be applied by adding external forces to the atoms near the top and bottom surfaces (usually within the cut-off radius of the interatomic potential). The external force is $\pm S\tau/N_{surf}$ where $N_{surf}$ is the number of atoms in the surface of area $S$, so that the total applied force is balanced between the upper and lower surfaces. We note that in case of an edge dislocation, the number of atoms in the top and bottom surfaces is not the same because of the extra half plane of the dislocation. Combinations of surface tractions parallel and perpendicular to the surfaces can be used to produce more complex stress tensors, including for instance, non-resolved components (Barvinschi et al. 2014).

After application of the initial displacement field, the atomic positions are relaxed through standard energy minimization algorithms, such as the conjugate gradient or damped dynamics, until typically the maximum atomic force component is below $10^{-4}$ eV/Å. As expected for FCC crystals, because of the existence of a stable stacking fault, the dislocation core splits in two partial dislocations separated by a stacking fault, the width of which depends on the stacking fault energy, the elastic constants, and the lattice parameter (Hirth and Lothe 1982). This is clearly seen in Fig. 2, where the atoms are colored according to their potential energy deviation from the cohesive energy of the perfect crystal. If the atom index is noted $i$ and $E_i^0$ is the reference energy, then with $E_i$ the potential energy of atom $i$ in the deformed cell, the potential energy deviation is $\Delta_i = E_i - E_i^0$. In the case where no impurity is present in the crystal, $E_i^0$ is simply the crystal cohesive energy per atom, $E_{coh}$.

In order to avoid spurious interactions with the periodic images, the dimension along the glide direction must be large enough (Proville and Patinet 2010). By increasing gradually the applied stress and relaxing the atomic positions, the dislocation starts to glide when the stress reaches a threshold, which corresponds to the dislocation Peierls stress. In the case where an impurity is situated close to the glide plane of the dislocation (as shown in Fig. 3), the stress threshold is larger than the Peierls stress, which is representative of solid solution hardening (Proville and Patinet 2010).

## 3 Kink-Pair Mechanism

When the applied resolved shear stress is below the Peierls stress, the dislocation has stable positions in its glide plane and moves at low temperature from one stable position to the next by a thermally activated process. The stable positions, called Peierls valleys, usually correspond to the dislocation aligned along a low-index crystallographic direction. The archetype is the screw dislocation in BCC metals (Dorn and Rajnak 1964; Guyot and Dorn 1967; Caillard and Martin 2003), where the dislocation line aligns with the dense $\langle 111 \rangle$ direction of its Burgers vector. The dislocation is then in a minimum-energy configuration separated from the next
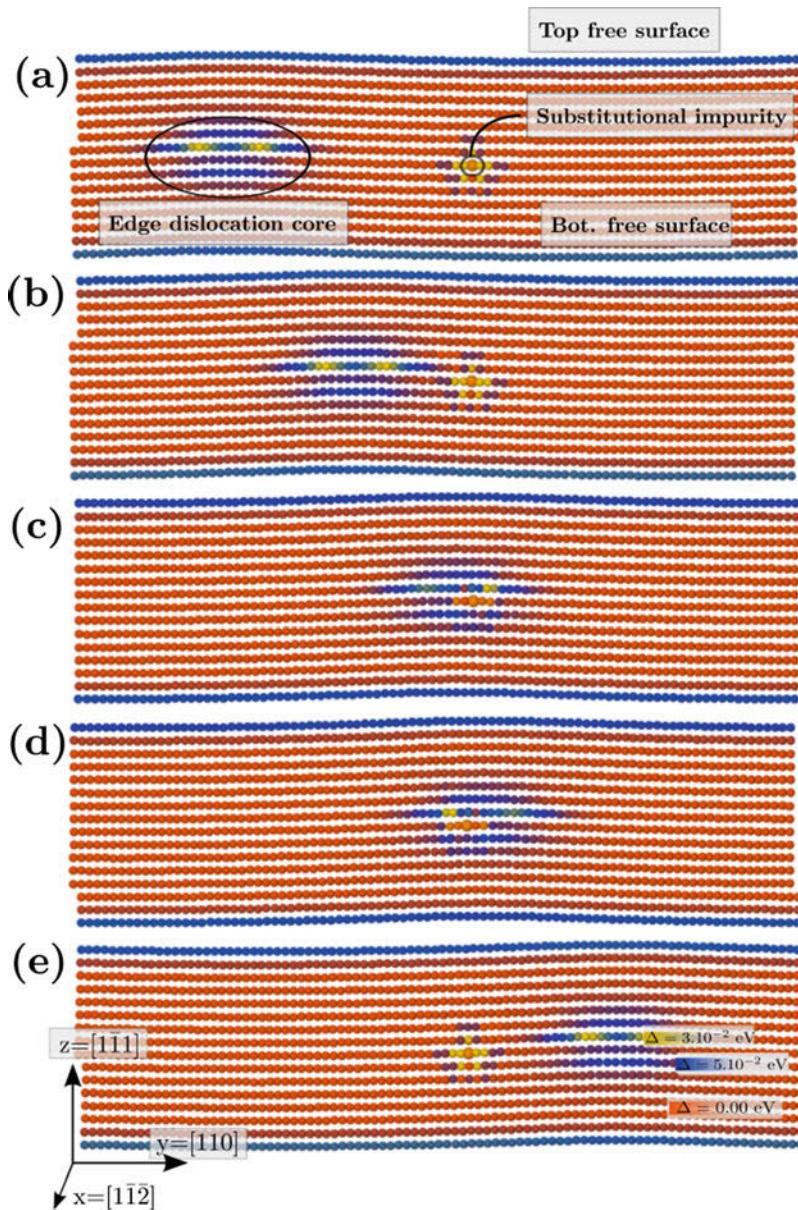
**Fig. 3** Snapshots of the crossing process for an $a/2[110](1\bar{1}1)$ edge dislocation in a FCC Al crystal with a single substitutional Mg impurity situated in the first crystal plane beneath the dislocation glide plane. Atomic colors are fixed according to the potential energy deviations from the perfect crystal cohesive energy
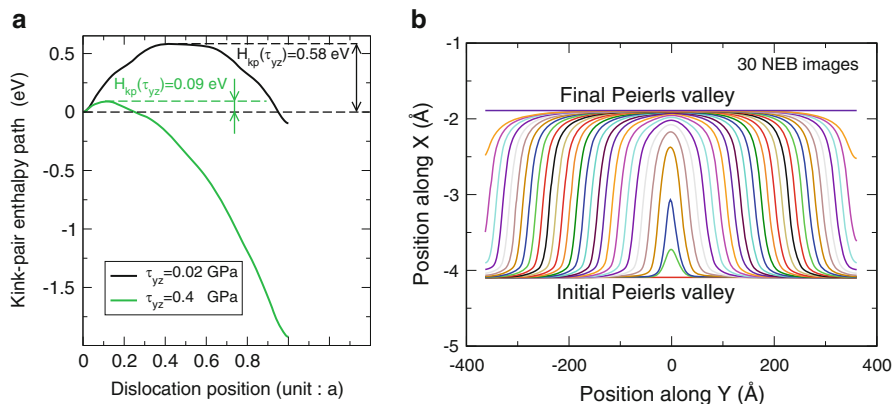
**Fig. 4** (**a**) Activation enthalpy for kink-pair formation on a screw dislocation in an EAM model of BCC $\alpha$-iron (Gordon et al. 2011) computed with nudged elastic band (NEB) method (Henkelman et al. 2000). (**b**) Profile of the dislocation in different NEB images showing the nucleation and expansion of a kink pair

equilibrium configuration by an energy barrier, called the Peierls barrier. To glide, the dislocation needs to jump to the next Peierls valley by a thermally activated process, which involves the nucleation of a short and localized segment in the next Peierls valley. This segment is bordered by two kinks, i.e., short dislocation segments that connect the dislocation between Peierls valleys. Once nucleated, this kink pair expands under the effect of the external stress, thereby progressively transferring the rest of the dislocation in the new Peierls valley.

This process can be computed at the atomic scale as illustrated in Fig. 4. Two configurations are generated with the dislocation in successive Peierls valleys. The nudged elastic band (NEB) method (Henkelman et al. 2000) is then used to identify the minimum energy path (MEP) between these configurations. The NEB method requires an initial path which must break the translational symmetry along the dislocation line to relax toward a MEP that reproduces the nucleation and propagation of a kink pair. Otherwise, the NEB method will converge to a high-energy path where the dislocation remains straight during the transition between Peierls valleys. To construct the intermediate states along the initial NEB path, atomic coordinates from the initial and final configurations are combined along the dislocation line. More precisely, the atomic coordinates are taken from the initial configuration except in a central region, whose width increases with the NEB image index. In this central region, the atomic coordinates are taken from the final state. A large number of NEB images allows to determine precisely the MEP. Examples under two different applied stresses are shown in Fig. 4a for a screw dislocation in a BCC iron model. The successive NEB images correspond to dislocation profiles, which are presented in Fig. 4b. As can be seen, the dislocation changes Peierls valley by the expected kink-pair mechanism. These dislocation profiles were obtained from the atomistic configurations by weighting the atomic positions with the atomic potential energy deviations, $\Delta_i$:

$$X_D = \frac{1}{\sum_i \Delta_i} \sum_i x_i \Delta_i \qquad (1)$$

$$Y_D = \frac{1}{\sum_i \Delta_i} \sum_i y_i \Delta_i$$

$$Z_D = \frac{1}{\sum_i \Delta_i} \sum_i z_i \Delta_i .$$

For a screw dislocation with a Burgers vector along Y axis, the dislocation position along the $X$ and $Z$ axis is expressed by $(X_D, Z_D)$, while for an edge dislocation with same Burgers vector, the position along the $Y$ and $Z$ axis is expressed by $(Y_D, Z_D)$. Then to determine the position of a dislocation segment, the computation of the position is performed within slices of crystal perpendicular to the dislocation line (see Landeiro Dos Reis et al. (2017) for more details). The dislocation position can be obtained from other considerations detailed in the literature (see, for instance, (Gröger and Vitek 2015; Dezerald et al. 2016) and Refs. therein), but the main advantages of the above definition are the following: (i) its simplicity when interatomic potentials are used and (ii) it can be generalized to other crystal defects, such as vacancies and interstitial atoms, using all three equations above to fully determine the defect position in the three dimensions of space. In the case of a long dislocation as in Fig. 4, the dislocation profile was obtained by computing the dislocation position in slices in direction $Y$ (the dislocation line direction) of width equal to the Burgers vector.

In case of a BCC screw dislocation, the migration barrier of the kinks along the dislocation line (called secondary Peierls barrier) is negligible, and the kink motion under stress is not thermally activated. The controlling process is then the nucleation of the initial unstable kink pair. The situation is different for instance in semiconductors, where the secondary Peierls valley is large and the diffusion of the kinks along the dislocation may be the controlling process (Hirth and Lothe 1982), depending on the temperature and the applied stress. We will however consider here only the case of BCC screw dislocations. At higher temperatures, in the so-called athermal regime, the thermal energy becomes large enough that dislocation motion is no longer thermally activated and resembles that observed in FCC metals.

As seen in Fig. 4a, the enthalpy barrier against kink-pair formation decreases with increasing applied shear stress due to the contribution of the work of the applied stress, which tilts the potential energy landscape. The stress at which the barrier disappears corresponds to the Peierls stress, above which dislocation motion is no longer thermally activated. This stress is obviously the same as the critical stress above which a stress-controlled enthalpy minimization does not find an equilibrium configuration but glides continuously in the simulation cell during the enthalpy minimization.

It has long been recognized that measurements of the Peierls stress from atomistic models systematically overestimate experimental data. This discrepancy was first reported by Basinski et al. (1971) in their early calculations in BCC

sodium modeled with an interatomic pair potential. But the same effect has since been repeatedly reported in BCC metals using more advanced energetic models, including EAM potentials (Wen and Ngan 2000; Chaussidon et al. 2006; Gordon et al. 2010), bond-order potentials (Gröger and Vitek 2008; Mrovec et al. 2011; Chen et al. 2013), and even ab initio density functional theory (DFT) (Woodward and Rao 2002; Ventelon and Willaime 2007; Ventelon et al. 2013; Weinberger et al. 2013). It was shown (Proville et al. 2012; Barvinschi et al. 2014) that the difference between experimental and simulated Peierls stresses is in part due to a quantum effect arising from the zero-point motion of the atoms near the dislocation core. This macroscopic quantum effect had so far been systematically discarded in atomistic simulations. However, such effect is expected even in heavy metals like Fe, because the Debye temperature, below which quantum effects in the vibrational modes of the system appear, is high (470 K in Fe) and the experiments to measure the Peierls stress were performed at very low temperatures, 4 K or less (Kuramoto et al. 1979; Brunner and Diehl 1992), well below the Debye temperature.

In contrast with glide above the Peierls stress, the thermally activated glide of dislocations below the Peierls stress cannot be studied using MD because the rate of kink-pair nucleation decreases exponentially rapidly when the stress decreases below the Peierls stress and the dislocation becomes immobile on MD timescales. On the other hand, the thermally activated nucleation of kink pairs can be modeled using the harmonic transition state theory (TST) (Benderskii et al. 1994). Moreover, this theory, which is the subject of the next section, can be written in a quantum form which allows to include zero-point energy corrections at low temperatures (Proville et al. 2012). Unlike in classical mechanics, quantum systems constantly fluctuate around their lowest energy state due to the Heisenberg uncertainty principle. The existence of non-zero minimum energy for a particle in a potential well $V(x) = m\omega_0^2/2x^2$ which is zero at the minimum position $x = 0$ can be demonstrated as follows. The Heisenberg uncertainty imposes that the fluctuations of the momentum $\delta p > \hbar/2\delta x$ which leads to a total energy $E = p^2/2m + V(x) > [(\hbar/2\delta x)^2/2m + V(\delta x)]$ which once minimized against $\delta x$ yields $\delta x^2 = \hbar/2m\omega_0$ and therefore $E > \hbar\omega_0/2$. Even though in a classical scheme the total energy is zero when both $x = 0$ and $p = 0$, the quantum fluctuations impose that the total energy cannot be less than $\hbar\omega_0/2$ which is called the zero-point energy. Within Einstein's model of crystals (Ashcroft and Mermin 1976), $\omega_0$ represents the Debye frequency, of the order of $\omega_0 \approx 60$ THz in Fe. This yields a zero-point energy per atom of 227 K, below which quantum fluctuations cannot be ignored.

## 4    Harmonic Transition State Theory

We consider a BCC crystal at low temperature where dislocation glide is thermally activated. Assuming that the plastic deformation is controlled by the glide of screw dislocations and not by their interactions with the microstructure, the mobility law equals the average velocity, and the plastic strain rate is given by Orowan law:

$$\dot{\varepsilon}_p = \rho_m b v_d \qquad (2)$$

where $\rho_m$ is the density of mobile screw dislocations, $b$ the Burgers vector, and $v_d$ the screw dislocation mobility law. Dislocation glide being controlled by the formation of kink pairs between nearest Peierls valleys, the dislocation velocity is expressed as $v_d = d\Gamma$ where $d$ is the distance between Peierls valleys and $\Gamma$ the kink-pair nucleation rate. According to the classical TST (Miller 1975), the rate $\Gamma$ for a three-dimensional system made of $N$ atoms of equal mass $m$ is expressed as an integral over phase space:

$$\Gamma = Z_r^{-1} \int \prod_{i=1}^{3N} \frac{dX_i dP_i}{h^{3N}} \delta((\mathbf{X} - \mathbf{X}_s).\mathbf{d}_s) \frac{|\mathbf{P}.\mathbf{d}_s|}{m} \theta(\mathbf{P}.\mathbf{d}_s) \exp\{[-\beta \mathscr{H}(\mathbf{P}, \mathbf{X})]\}, \qquad (3)$$

where thick symbols stand for 3N-dimensional vectors, $\beta \equiv 1/k_B T$, and $h$ is Planck's constant. $\mathbf{X}$ is the configuration position with $\mathbf{X}_s$ the transition state, i.e., the unstable dislocation configuration between Peierls valleys. $\mathbf{d}_s$ is the corresponding unstable eigenmode, i.e., the direction in configuration space from $\mathbf{X}_s$ where the energy has a negative curvature. This direction is perpendicular to the dividing surface, which separates at $\mathbf{X}_s$ the basins of attraction of the initial and final configurations, i.e., the dislocation in two successive Peierls valleys. The Dirac function term, $\delta((\mathbf{X} - \mathbf{X}_s).\mathbf{d}_s)$, therefore limits the configuration integral to the dividing surface. $\mathbf{P}$ is the momentum, and the Heaviside function term $\theta(\mathbf{P}.\mathbf{d}_s)$ ensures that the integration in momentum space is performed in the half-space toward the product, i.e., the dislocation in the next Peierls valley. $\mathscr{H}(\mathbf{P}, \mathbf{X})$ is the Hamiltonian of the system to which the work of applied stress is properly substracted. Finally, $Z_r$ is the reactant partition function, that is, the harmonic partition function of the crystal with the dislocation at the bottom in its Peierls valley. The three free translational modes of the system have no contribution to the rate $\Gamma$ since they equally contribute to $Z_r$ and to the numerator in Eq. 3. They are omitted in the following.

To introduce the harmonic approximation, we use a normal mode representation $(\mathbf{q}, \mathbf{p})$ computed at the saddle state, instead of the cartesian coordinates $(\mathbf{X}, \mathbf{P})$. The unstable mode is placed at the end of the mode list, i.e., $s = 3N - 3$. If we note $\lambda_\nu$ the eigenvalue of mode $\nu$, the kink-pair nucleation rate is then written as

$$\Gamma = Z_r^{-1} \int \frac{dq_s dp_s}{h} \frac{p_s}{m} \theta(p_s)\delta(q_s) \prod_{\nu=1}^{3N-4} \frac{dq_\nu dp_\nu}{h^{3N-4}}$$

$$\exp\left\{\left[-\beta\left(H_{kp} + \sum_{\nu=1}^{3N-3}[\frac{p_\nu^2}{2m} + \lambda_\nu \frac{q_\nu^2}{2}]\right)\right]\right\}, \qquad (4)$$

where $H_{kp}$ is the activation enthalpy for kink-pair formation (the enthalpy difference between saddle and initial states), function of the applied shear stress $\tau_{yz}$.

In BCC crystals, in the stress regime considered here, the secondary Peierls barrier is absent. As a result, there is no resistance for kink motion along the dislocation line, resulting in a Goldstone mode of zero eigenvalue. The latter, assumed to correspond to $\nu = 1$, can be integrated out (it corresponds to the Gaussian integration of a kinetic energy term $1/h \int dq dp \exp(-\beta p^2/2m) = L_d \sqrt{2\pi mkT}/h$), yielding:

$$\Gamma = \frac{L_d \sqrt{2\pi mkT}}{h Z_r} \int \frac{dq_s dp_s}{h} \frac{p_s}{m} \theta(p_s)\delta(q_s) \prod_{\nu=2}^{3N-4} \frac{dq_\nu dp_\nu}{h^{3N-5}}$$

$$\exp\left\{\left[-\beta\left(H_{kp} + \sum_{\nu=2}^{3N-4}[\frac{p_\nu^2}{2m} + \lambda_\nu \frac{q_\nu^2}{2}]\right)\right]\right\}. \tag{5}$$

Note that the dislocation length $L_d$ now appears explicitly in the nucleation rate as a result of the integration of the Goldstone mode. This length dependence of the dislocation velocity has been measured experimentally by in situ TEM (Caillard 2010).

At low temperature, quantum mechanics must be accounted for. Since the normal modes are independent of one another, each stable mode (for $\nu = 2$ to $3N-4$) yields in Eq. 5 the partition function of a quantum harmonic oscillator $[2\sinh\left(\frac{\hbar\beta}{2}\omega_\nu\right)]^{-1}$, where $\omega_\nu = \sqrt{\lambda_\nu/m}$ is the angular frequency of mode $\nu$. Similarly, the partition function in the stable reactant state $Z_r$ can be treated harmonically, yielding $Z_r = \prod_{\mu=1}^{3N-3}[2\sinh\left(\frac{\hbar\beta}{2}\Omega_\mu\right)]^{-1}$, with $\{\Omega_\mu\}$ the angular frequencies associated with the reactant state. The nucleation rate is then expressed as

$$\Gamma = \frac{L_d\sqrt{2\pi mkT} \prod_{\mu=1}^{3N-3}[2\sinh\left(\frac{\hbar\beta}{2}\Omega_\mu\right)]}{h \prod_{\nu=2}^{3N-4}[2\sinh\left(\frac{\hbar\beta}{2}\omega_\nu\right)]} I, \tag{6}$$

where the integral $I$ is defined by

$$I = \int \frac{dq_s dp_s}{h} \frac{p_s}{m} \theta(p_s)\delta(q_s) \exp\left\{\left[-\beta\left(\frac{p_s^2}{2m} + H_{kp} + \lambda_s \frac{q_s^2}{2}\right)\right]\right\}. \tag{7}$$

This integral over the $q_s$ coordinate of the unstable mode is replaced (Benderskii et al. 1994) by the crossing rate of the parabolic barrier $H_{kp} + \lambda_s \frac{q_s^2}{2}$, which yields

$$I = \frac{1}{h} \int dE \; W(E) \exp\{(-\beta E)\}, \tag{8}$$

where $W$ the transmission coefficient for a parabolic barrier (Landau and Lifshitz 1981), given by $W(E) = [1 + \exp\{(2\pi(H_{kp} - E)/\hbar\omega_s)\}]^{-1}$ and $\omega_s = \sqrt{-\lambda_s/m}$.

After simplication, the integral becomes:

$$I = \frac{kT}{h} \frac{\hbar\beta\omega_s/2}{\sin(\hbar\beta\omega_s/2)} e^{-\beta H_{kp}}. \tag{9}$$

Within the harmonic TST, the rate of kink-pair formation is therefore expressed as

$$\Gamma = L_d \sqrt{\frac{2\pi m}{kT}} \frac{\hbar\beta\omega_s/2}{\sin(\hbar\beta\omega_s/2)} \left(\frac{kT}{h}\right)^2 \frac{\prod_{\mu=1}^{3N-3}[2\sinh\left(\frac{\hbar\beta}{2}\Omega_\mu\right)]}{\prod_{\nu=2}^{3N-4}[2\sinh\left(\frac{\hbar\beta}{2}\omega_\nu\right)]} e^{-\beta H_{kp}}. \tag{10}$$

Combining this equation with Eq. 2, we can therefore express the plastic strain rate as a quantum thermally activated Orowan law. We note that at high temperature when $\beta = 1/k_B T \to 0$, the sin and sinh terms become equivalent to their arguments and Eq. 2 can be simplified. The well-known TST expression characteristic of the classical regime is then recovered (Vineyard 1957):

$$\Gamma_C = L_d \sqrt{\frac{2\pi m}{kT}} \frac{\prod_{\mu=1}^{3N-3} \Omega_\mu}{\prod_{\nu=2}^{3N-4} \omega_\nu} e^{-\beta H_{kp}}, \tag{11}$$

where the pre-exponential factor involves the Goldstone mode and the ratio of the product of the eigenfrequencies in the initial and activated states. In experimental conditions, the plastic strain rate $\dot{\varepsilon}_p$ is constant and so is the density of mobile dislocations, $\rho_d$, at least to a first-order approximation. For a given temperature, Orowan equation (Eq. 2) is therefore solved self-consistently, finding the appropriate applied stress $\tau_{yz}$, which sets the saddle state and thus sets $H_{kp}$, $\omega_s$, $\{\omega_\nu\}$ and $\{\Omega_\mu\}$, and produces the kink-pair nucleation rate, which satisfies Orowan's equation.

It is well-known (Benderskii et al. 1994; Gillan 1987) that Eq. 9 diverges at a critical temperature $T_{\text{tun}} = \hbar\omega_s/2\pi k$, which marks the onset of deep quantum tunneling. The above harmonic theory is thus valid only above $T_{\text{tun}}$, but for iron, this temperature is small, typically below 10 K. Below $T_{\text{tun}}$, nonlinear effects must be accounted for, which is the subject of the instanton theory developed by Miller (1975). The most important contribution of quantum mechanics comes from the crystal partition functions that appear in the pre-exponential factor in Eq. 10. This quantum correction has some similarities with the heat capacity deviation from the Dulong-Petit law (see textbooks as Ashcroft and Mermin 1976).

## 5    Numerical Implementation

Computing Eq. 10 requires to evaluate the eigenfrequencies of the system and therefore requires to diagonalize the Hessian matrix. The latter can be obtained analytically if the interatomic potential is itself analytical. Otherwise, the Hessian

matrix coefficients are obtained through the calculation of the first derivative of the forces under a finite displacement of each atom in the three directions of space. When atom $i$ is displaced in direction $d_i$, the force field must be computed for each neighbor $j$, in each direction $d_j$. Denoting by $F_{d_j,j}$, the force exerted on atom $j$ in direction $d_j$, and by $X_{d_i,i}$, the displacement of atom $i$ in direction $d_i$, the coefficient of the Hessian matrix situated at row $i + N_{at}(d_i - 1)$ and column $j + N_{at}(d_j - 1)$ is given by

$$\mathcal{H}_{i+N_{at}(d_i-1),\, j+N_{at}(d_j-1)} = -\frac{1}{\sqrt{m_i m_j}} \frac{\delta F_{d_j,j}}{\delta X_{d_i,i}}, \tag{12}$$

where $m_i$ and $m_j$ are the masses of atoms $i$ and $j$, respectively. Displacements $X_{d_i,i}$ are typically one hundredth of an Angström. The exact diagonalization of the matrix $\mathcal{H}$ can be performed using the linear algebra package LAPACK (Anderson et al. 1999). Different diagonalization methods can be used for small simulation cells, with a number of atoms typically $N_{at} < 10,000$. However treating larger systems requires the use of a distributed memory algorithm because the Hessian matrix size overpasses the capabilities of serial processors. In this case, we employed ScaLAPACK parallel routines.

The result of the calculations is shown in Fig. 5 for two EAM models of $\alpha$-iron. We compare here the quantum (Eq. 10) and classical (Eq. 11) TST expressions. We see that in both cases, the classical expression strongly overestimates experimental
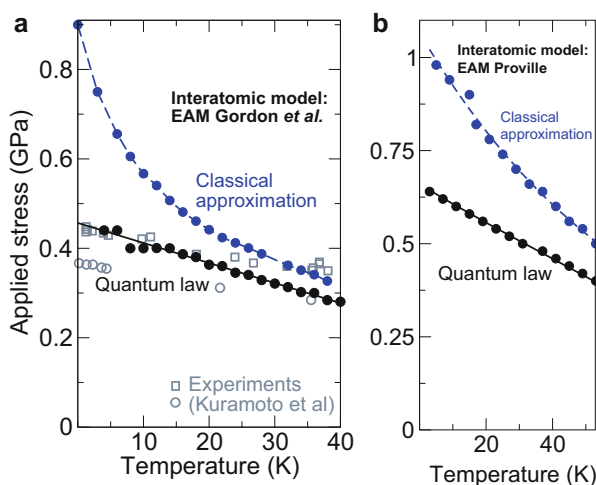


**Fig. 5** (**a**) Comparison of the temperature dependence of the flow stress predicted through the quantum Orowan law (solid curve), its classical approximation (dashed curve) and experimental data obtained in iron (the open circles and squares refer to two different orientations of the traction axis) (Kuramoto et al. 1979). The calculations were performed with the EAM potential for $\alpha$-iron developed by Gordon et al. (2011). (**b**) The same computations as in (**a**) with a different energetic model for iron (Proville et al. 2012)

data because of a rapid increase of the flow stress at low temperatures, while the quantum law does not show this rapid increase and is much closer to the experimental data, particularly in the case of the EAM potential developed by Gordon et al. (2011).

To evaluate the importance of the quantum correction, the zero-point energy of the crystal can be computed along the MEP from the Hessian diagonalization of each NEB image. If the $3N$ eigenvalues of $\mathscr{H}$ are denoted by $\lambda_u$, the zero-point energy of the system is evaluated through the harmonic approximation as

$$U_{ZP} = \frac{\hbar}{2} \sum_{u=1}^{3N} \sqrt{\lambda_u}, \tag{13}$$

where the unstable mode for which $\lambda_u < 0$ is excluded from the summation. The amplitude of the associated imaginary frequencies remains negligible in comparison to the variations of the zero-point energy along the MEP. The latter is reported in Fig. 6a for a straight screw dislocation as a function of its position in the crystal.

According to the quantum harmonic theory employed here, one can expect that the crystal quantum fluctuations are large enough to reduce the effect of the Peierls barriers the dislocations have to overcome. Figure 6a shows the case of a screw dislocation in $\alpha$-Fe. We see that the variation of the zero-point energy is opposite to that of the Peierls barrier and is approximately half its amplitude with the present interatomic potential (Proville et al. 2012). In absence of thermal
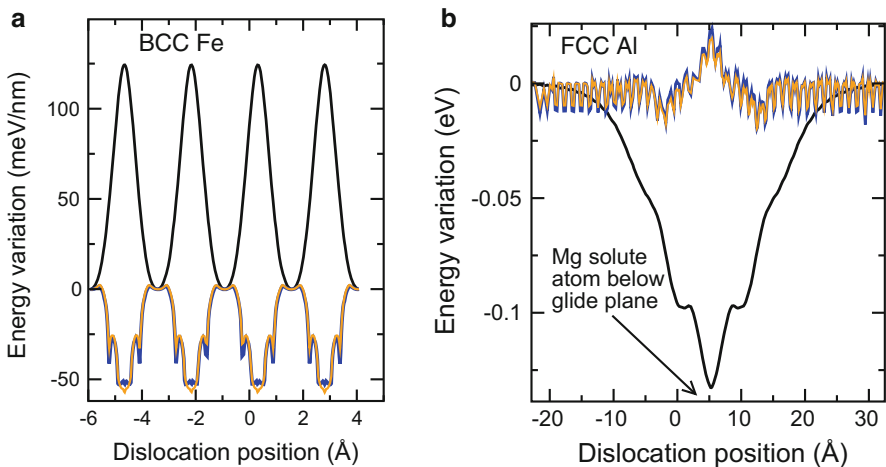


**Fig. 6** Comparison between the potential and zero-point energy variations during the glide of a straight screw dislocation in BCC $\alpha$-Fe (**a**) and of an edge dislocation in FCC Al in presence of a Mg solute atom in the plane just below the glide plane (**b**). The energies were computed along the minimum potential energy path obtained with the NEB method. The potential energy is shown in black, and the zero-point energy is computed from the exact diagonalization of the Hessian matrix (blue) and from an Einstein approximation (orange)

effects, the decrease of the Peierls stress by quantum fluctuations can therefore be understood from the consideration that the total energy of the system is composed of the sum of (i) the potential energy which is computed from the EAM interaction potential and (ii) the zero-point energy, which is itself the result of a combination between kinetic and vibrational energies. The zero-point energy variation has been computed in various FCC crystals following the same method (Landeiro Dos Reis et al. 2017). Figure 6b shows the example of an edge dislocation crossing an Mg solute in a FCC crystal of Al (see Fig. 3a–e). The impurity is located below the glide plane in a region of attraction and therefore induces a negative variation of the potential energy. We see in this example that the variation of the zero-point energy is again opposite to the potential energy and therefore decreases the attraction between the dislocation and the Mg solute, although the effect is not as marked as for the BCC screw dislocation. In all cases considered so far, the variation of the zero-point energy was opposite to the potential energy and therefore systematically decreases the energy barrier.

To gain some insight into the zero-point energy variation along the MEP of dislocations, we tested the approximation, which consists in computing the eigenfrequencies from the on-site frequencies of each atom with the other atoms frozen, that is, from the diagonal elements of the Hessian matrix only. The solid is then modeled as an assembly of independent oscillators, which corresponds to an Einstein approximation. The corresponding zero-point energy variation is shown in orange in Fig. 6 and closely follows the full calculation in both BCC and FCC crystals (Landeiro Dos Reis et al. 2017). This very good agreement between the Einstein approximation and the exact zero-point energy calculation shows that the main quantum contribution comes from vibrational modes with short wavelengths, since the acoustic modes are not included in the Einstein approximation.

## 6    Conclusion

The thermally activated mobility of dislocations can be studied at the atomic scale in model crystals with the help of the transition state theory to predict the dislocation jump rate. All the parameters of the model can be derived from the atomic scale, demonstrating the feasibility of a multi-scale modeling with no adjustable parameter. Comparison with experimental deformation tests in BCC Fe shows that the predictions established from the present method are satisfactory, provided that the quantum effect due to zero-point energy fluctuations is taken into account at low temperatures. To that purpose, we have employed a harmonic approximation. Both the TST and the harmonic approximation are the purpose of constant improvements (Mills et al. 1995), but these developments imply in general a much heavier computational load, which is difficult to apply to systems including thousands of atoms as is typical in simulation cells with dislocations. In pure BCC crystals where the Peierls landscape is the main barrier impeding screw dislocation glide, we have found that the zero-point energy variation is negative and therefore eases the motion of the dislocation. In FCC solid solutions, the interaction between

dislocations and solute atoms is the main barrier and can be either attractive or repulsive. However, the zero-point energy variation is systematically opposite to that of the potential energy and therefore always eases dislocation glide. The magnitude of the effect however depends on the case considered. Such a systematic observation through atomic simulations allows us to expect that the softening effect due to zero-point energy fluctuations is ubiquitous, although such a conclusion should require more theoretical investigations.

Based on the developments presented here, we believe that the present atomic-scale solid-state theory is getting close to establish predictions conforming to the experimental deformation tests performed in model materials as pure BCC crystals. However many challenges still remain to be faced, and questions remain to be addressed, as, for instance:

(i) We have used the simplest approximation (harmonic TST) but more accurate rate theories, which in particular include anharmonicity, should be tested (Saroukhani and Warner 2017).

(ii) Can we obtain more quantitative estimates of the zero-point energy using ab initio calculations? The accuracy of Einstein approximation allows to significantly reduce the computational cost, but computing force derivatives requires a precision higher than usual.

(iii) Ab initio calculations are still too computationally demanding to model kinked dislocations, but we should popularize the use of higher-scale models, like the line tension model (Dezerald et al. 2015), to predict dislocation mobility based on quantitative ab initio data.

(iv) In the particular case of BCC crystals, non-Schmid effects are known to significantly affect dislocation mobility (Dezerald et al. 2016), and we should in particular include the effect of non-glide components of the stress tensor (Barvinschi et al. 2014; Cereceda et al. 2016), which are unavoidable in experimental conditions of uniaxial tensile tests.

(v) The case of high-secondary Peierls valleys as in semiconductors or ceramics should also be investigated.

# References

Alshits VI (1992) Elastic strain fields and dislocation mobility. North-Holland, Amsterdam

Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, Sorensen D (1999) LAPACK users' guide, 3rd edn. Society for Industrial and Applied Mathematics, Philadelphia

Ashcroft NW, Mermin ND (1976) Solid state physics. Saunders College Publishing, Philadelphia

Bacon DJ, Osetsky YN, Rodney D (2008) In: Hirth J, Kubin L (eds) Dislocations in solids, Dislocation-Obstacle Interactions at the Atomic Level. Elsevier, Amsterdam

Barvinschi B, Proville L, Rodney D (2014) Quantum Peierls stress of straight and kinked dislocations and effect of non-glide stresses. Model Simul Mater Sci Eng 22:025006

Basinski ZS, Duesbery MS, Taylor R (1971) Influence of shear stress on screw dislocations in a model sodium lattice. Can J Phys 49:2160–2180

Benderskii V, Makarov D, Wight C (1994) Chemical dynamics at low temperature. Wiley-Interscience, New York

Bhate N, Clifton R, Phillips R (2002) Atomistic simulations of the motion of an edge dislocation in aluminum using the embedded atom method. In: AIP conference proceedings, vol 620. American Institute of Physics, Atlanta, Georgia pp 339–342

Brunner D, Diehl J (1992) Extension of measurements of the tensile flow stress of high-purity α-iron single crystals to very low temperatures. Z Metalkd 83:828

Bulatov VV, Cai W (2006) Computer simulations of dislocations. Oxford University Press, New York

Cai W, Bulatov VV (2004) Mobility laws in dislocation dynamics simulations. Mater Sci Eng A 387:277–281

Caillard D (2010) Kinetics of dislocations in pure Fe. Part II. In situ straining experiments at low temperature. Acta Mater 58:3504–3515

Caillard D, Martin JL (2003) Thermally activated mechanisms in crystal plasticity. Pergamon, Amsterdam

Cereceda D, Diehl M, Roters F, Raabe D, Perlado JM, Marian J (2016) Unraveling the temperature dependence of the yield strength in single-crystal tungsten using atomistically-informed crystal plasticity calculations. Int J Plast 78:242–265

Chaussidon J, Fivel M, Rodney D (2006) The glide of screw dislocations in BCC Fe: atomistic static and dynamic simulations. Acta Mater 54:3407

Chen Z, Mrovec M, Gumbsch P (2013) Atomistic aspects of screw dislocation behavior in α-iron and the derivation of microscopic yield criterion. Model Simul Mater Sci Eng 21:055023

Dezerald L, Proville L, Ventelon L, Willaime F, Rodney D (2015) First-principles prediction of kink-pair activation enthalpy on screw dislocations in BCC transition metals: V, Nb, Ta, Mo, W, and Fe. Phys Rev B 91:094105

Dezerald L, Rodney D, Clouet E, Ventelon L, Willaime F (2016) Plastic anisotropy and dislocation trajectory in BCC metals. Nat Commun 7:11695

Dorn JE, Rajnak S (1964) Nucleation of kink pairs and the Peierls' mechanism of plastic deformation. Trans Metal Soc AIME 230:1052

Gilbert MR, Queyreau S, Marian J (2011) Stress and temperature dependence of screw dislocation mobility in α-Fe by molecular dynamics. Phys Rev B 84:174103

Gillan M (1987) Quantum-classical crossover of the transition rate in the damped double well. J Phys C 20:3621

Gordon PA, Neeraj T, Li Y, Li J (2010) Screw dislocation mobility in BCC metals: the role of the compact core on double-kink nucleation. Model Simul Mater Sci Eng 18:085008

Gordon PA, Neeraj T, Mendelev MI (2011) Screw dislocation mobility in BCC metals: a refined potential description for α-Fe. Philos Mag Lett 91:3931–3945

Gröger R, Vitek V (2008) Multiscale modeling of plastic deformation of molybdenum and tungsten. III. Effects of temperature and plastic strain rate. Acta Mater 56:5426–5439

Gröger R, Vitek V (2015) Determination of positions and curved transition pathways of screw dislocations in BCC crystals from atomic displacements. Mater Sci Eng A 643:203

Gumbsch P, Gao H (1999) Dislocations faster than the speed of sound. Science 283: 965–968

Guyot P, Dorn JE (1967) A critical review of the Peierls mechanism. Can J Phys 45:983

Henkelman G, Jóhannesson G, Jónsson H (2000) Methods for finding saddle points and minimum energy paths: theoretical methods in condensed phase chemistry, chap 10. In: Schwartz SD (ed) Progress in theoretical chemistry and physics, vol 5. Springer, Dordrecht (NL), pp 269–302

Hirth JP, Lothe J (1982) Theory of dislocations. Wiley, New York

Johnston WG, Gilman JJ (1959) Dislocation velocities, dislocation densities, and plastic flow in LiF crystals. J Appl Phys 30:129

Koizumi H, Kirchner HOK, Suzuki T (2002) Lattice wave emission from a moving dislocation. Phys Rev B 65:214104

Kubin L (2013) Dislocations, mesoscale simulations and plastic flow. Oxford University Press, Oxford

Kuramoto E, Aono Y, Kitajima K (1979) Thermally activated slip deformation between 0.7 and 77 K in high-purity iron single crystals. Philos Mag 39:717

Landau LD, Lifshitz EM (1981) Quantum mechanics non-relativistic theory. Elsevier Science Ltd, Amsterdam

Landeiro Dos Reis M, Choudhury A, Proville L (2017) Ubiquity of quantum zero-point fluctuations in dislocation glide. Phys Rev B 95:094103

Leibfried G (1950) Uber den einfluss thermisch angeregter schallwellen auf die plastische deformation. Z Phys 127:344

Marian J, Cai W, Bulatov VV (2004) Dynamic transitions from smooth to rough to twinning in dislocation motion. Nat Mater 3:158

Miller W (1975) Semiclassical limit of quantum mechanical transition state theory for nonseparable systems. J Chem Phys 62:1899

Mills G, Jónsson H, Schenter GK (1995) Reversible work transition state theory: application to dissociative adsorption of hydrogen. Surf Sci 324:305

Mordehai D, Ashkenazy Y, Kelson I, Makov G (2003) Dynamic properties of screw dislocations in Cu: a molecular dynamics study. Phys Rev B 67:24112

Mrovec M, Nguyen-Manh D, Elsasser C, Gumbsch P (2011) Magnetic bond-order potential for iron. Phys Rev Lett 106:246402

Nadgornyi E (1988) Dislocation dynamics and mechanical properties of crystals. Prog. Mater. Sci. 31:1–530

Nosenko V, Morfill G, Rosakis P (2011) Direct experimental measurement of the speed-stress relation for dislocations in a plasma crystal. Phys Rev Lett 106: 155002

Olmsted DL, Hector Jr LG, Curtin WA, Clifton RJ (2005) Atomistic simulations of dislocation mobility in Al, Ni and Al/Mg alloys. Model Simul Mater Sci Eng 13:371–388

Oren E, Yahel E, Makov G (2017) Dislocation kinematics: a molecular dynamics study in Cu. Model Simul Mater Sci Eng 25:025002

Osetsky YN, Bacon DJ (2003) An atomic-level model for studying the dynamics of edge dislocations in metals. Model Simul Mater Sci Eng 11:427

Patinet S, Proville L (2008) Depinning transition for a screw dislocation in a model solid solution. Phys Rev B 78:104109

Proville L, Patinet S (2010) Atomic-scale models for hardening in FCC solid solutions. Phys Rev B 82:054115

Proville L, Rodney D, Marinica MC (2012) Quantum effect on thermally activated glide of dislocations. Nat Mater 11:845–849

Rodney D (2004) Molecular dynamics simulation of screw dislocations interacting with interstitial frank loops in a model FCC crystal. Acta Mater 52:607–614

Rodney D, Martin G (2000) Dislocation pinning by glissile interstitial loops in a nickel crystal: a molecular-dynamics study. Phys Rev B 61:8714

Rodney D, Proville L (2009) Stress-dependent Peierls potential: influence on kink-pair activation. Phys Rev B 79:094108

Saroukhani S, Warner DH (2017) Investigating dislocation motion through a field of solutes with atomistic simulations and reaction rate theory. Acta Mater 128:77–86

Suzuki T, Takeuchi S, Yoshinaga H (1991) Dislocation dynamics and plasticity, Springer Series in Materials Science vol. 12. Springer, Berlin

Ventelon L, Willaime F (2007) Core structure and Peierls potential of screw dislocations in $\alpha$-fe from first principles: cluster versus dipole approaches. J Comput-Aided Mater Des 14: 85–94

Ventelon L, Willaime F, Clouet E, Rodney D (2013) Ab initio investigation of the Peierls potential of screw dislocations in BCC Fe and W. Acta Mater 61:3973

Vineyard GH (1957) Frequency factors and isotope effects in solid state rate processes. J Phys Chem Solids 3:121–127

Weinberger C, Tucker G, Foiles S (2013) Peierls potential of screw dislocations in BCC transition metals: predictions from density functional theory. Phys Rev B 87:054114

Wen M, Ngan AHW (2000) Atomistic simulation of kink-pairs of screw dislocations in body-centred cubic iron. Acta Mater 48:4255–4265

Woodward C, Rao SI (2002) Flexible ab initio boundary conditions: simulating isolated dislocations in BCC Mo and Ta. Phys Rev Lett 88:216402

# Dislocation Analysis Tool for Atomistic Simulations

# 72

Alexander Stukowski

## Contents

**Abstract**

Precise analysis and meaningful visualization of dislocation structures in molecular dynamics simulations are important steps toward physical insights. This chapter provides an introduction to the dislocation extraction algorithm (DXA), which is a computational method for identifying and quantifying dislocations in atomistic crystal models. It builds a bridge between the atomistic world of crystal defects and the discrete line picture of classical dislocation theory.

## 1    Introduction

Dislocations have two sides: On one hand, they are commonly viewed in dislocation theory as discrete line objects, which possess a characteristic topological charge – the *Burgers vector* – and which can glide through a crystal to produce plasticity or participate in various kinds of reactions. On the other hand, they constitute a partic-

A. Stukowski (✉)
Materials Modeling Division, Institute of Materials Science, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: stukowski@mm.tu-darmstadt.de
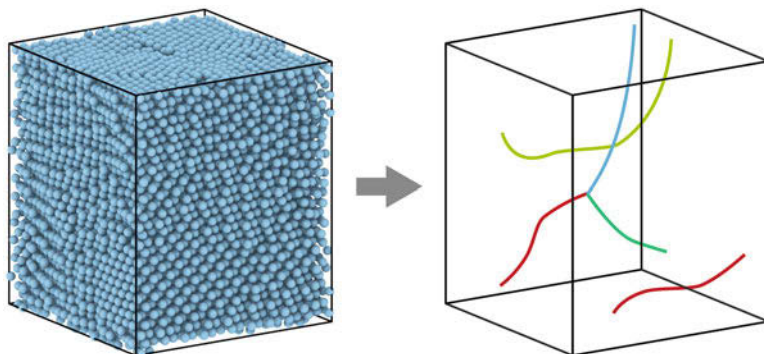
**Fig. 1** In this section a computer algorithm is introduced that can convert an atomistic crystal model (left) to a discrete line representation of the contained dislocation defects (right)

ular kind of irregularity in the otherwise regular arrangement of atoms in the crystal lattice. Boundaries of extra planes of atoms are what we call edge dislocations, and screw dislocations denote helical distortions of atomic planes. Thus, to fully grasp the phenomenon of crystal dislocations, we must consider – and somehow unify – two complementary pictures: the *explicit* one (discrete line objects) and the *implicit* one (glitches in the lattice arrangement). Both pictures are extensively used by different modeling strategies, with discrete dislocation dynamics (DDD) and molecular dynamics (MD) being the most prominent ones to represent the two dislocation descriptions. This section will introduce ways to unify these seemingly disparate pictures of the same physical phenomenon and to accomplish a conversion between them (Fig. 1). Given the atomic positions in a crystal containing dislocation defects, we want to reconstruct the geometry of the one-dimensional lines these dislocations can be described as mathematically. The computational method introduced here to solve this problem has practical relevance for the modelling of dislocations at multiple length scales as it builds a bridge between the atomistic world and the mesoscale and, at the same time, provides a powerful analysis tool for MD simulations that greatly helps to understand dislocation processes.

## 2 Burgers Circuit Method

As an introductory example for the connection between dislocation line theory and atomistic crystal defects, we take a look at the classical Burgers circuit construction (Frank 1951), which is the canonical method (Bulatov and Cai 2006) already proposed in the 1950s to discriminate dislocations from other crystal defects and to determine their Burgers vectors. In the formulation employed here, a Burgers circuit $C$ is a path in the dislocated crystal consisting of a sequence of atom-to-atom steps (line elements $\Delta\mathbf{x}$), as shown in Fig. 2a. The circuit is closed, thus $\sum_C \Delta\mathbf{x} = \mathbf{0}$.

We assume that there exists a mapping $\Delta\mathbf{x} \rightarrow \Delta\mathbf{x}'$ that translates each line element of the path to a corresponding image, $\Delta\mathbf{x}'$, in a perfect crystal
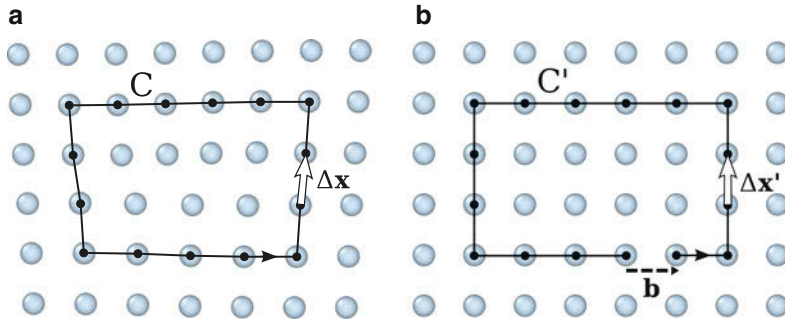
**Fig. 2** Burgers circuit method to identify a dislocation. A closed circuit around the dislocation is translated from (**a**) the dislocated crystal to (**b**) the perfect reference crystal. The closure failure **b** is called the Burgers vector of the dislocation

lattice (Fig. 2b). Summing these transformed line elements algebraically along the associated path, $C'$, gives the true Burgers vector of the dislocation enclosed by $C$:

$$\mathbf{b} = -\sum_{C'} \Delta\mathbf{x}'. \tag{1}$$

The Burgers vector **b** is the closure failure of the path after transferring it to the perfect reference crystal.

   Note that the Burgers circuit procedure is typically performed by hand to analyze two-dimensional crystal images obtained from high-resolution microscopy or atomistic computer simulations. Human intuition and cognitive capabilities are required to spot irregularities in the crystal lattice that are potential dislocation defects and to apply the Burgers circuit test to them. Automating these steps poses a challenge when developing a computational dislocation identification method for three-dimensional atomistic crystal models.

   The Burgers circuit procedure represented by Eq. 1 above is the discrete analogue of an equation used in continuum mechanics to define the Burgers vector of a Volterra dislocation:

$$\mathbf{b} = -\int_C (\mathbf{F}^{\mathbf{e}})^{-1}\mathrm{d}\mathbf{x}. \tag{2}$$

Here, $C$ denotes any contour enclosing the mathematical dislocation line, and $(\mathbf{F}^{\mathbf{e}})^{-1}$ denotes the inverse of the elastic deformation gradient. This second-rank tensor acts on the infinitesimal line element $\mathrm{d}\mathbf{x}$ and transforms it from the dislocated crystal configuration to an ideal, elastically unstrained reference configuration. This mapping is analogous to the explicit translation of atomic steps we did in the discrete formulation of the Burgers circuit procedure.

   Notably, the resulting vector **b** stays the same if we change the original circuit $C$, as long as it still encloses the same dislocation. On the other hand, if $\mathbf{b} = \mathbf{0}$, we
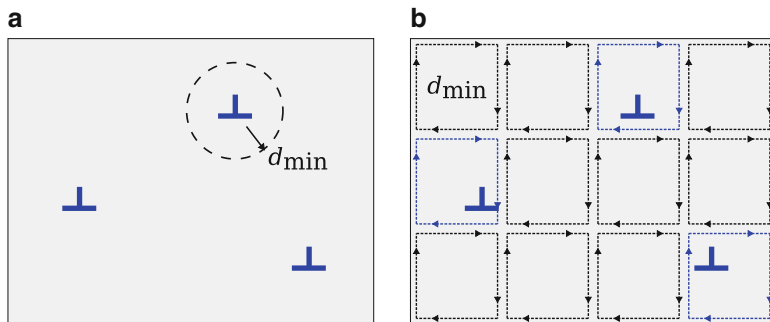
**Fig. 3** Schematic depiction of the dislocation finding approach described in the text. (**a**) The given domain contains a set of dislocations with unknown positions. The parameter $d_{\min}$ denotes the lower bound for the separation distance between dislocations, which is on the order of one atomic lattice spacing and corresponds to the dislocation core diameter. (**b**) The domain is tessellated by a grid of Burgers circuits of diameter $d_{\min}$. Circuits highlighted in blue exhibit a closure failure and are marked as containing a dislocation

know that the Burgers circuit did not enclose any defect with dislocation character. Here, however, we are deliberately ignoring the possibility that the circuit encloses multiple dislocations whose Burgers vectors cancel. One may thus ask, in the absence of a priori knowledge of the spatial distribution of dislocations in a given crystal, how can we – or rather a computer algorithm – construct the circuit $C$ such that it encloses exactly one of the dislocations?

This general situation is depicted schematically in Fig. 3a: A set of dislocations with unknown positions which we would like to determine is distributed across a given continuum domain. It is safe to assume that any distribution of dislocations is such that one can specify a lower-bound $d_{\min}$ for the separation distance between any two distinct dislocations. In reality, this lower bound is given by the interatomic spacing in the crystal lattice, because that is also the minimum distance when the cores of two nearby dislocations necessarily start to overlap; hence they can no longer be treated as distinct defects. As shown in Fig. 3b, it is possible to construct a large number of non-overlapping circuits, each having size $d_{\min}$, to completely cover the entire domain. By virtue of our construction, each circuit can contain at most one dislocation, and we effectively excluded the possibility of "missing" dislocations. The Burgers circuit test tells us which of the circuits contain a dislocation, and since their diameters are small, we can pinpoint the dislocations' positions with great precision (on the order of $d_{\min}$) using this method.

## 3 Simple Algorithm for Finding Dislocations in Atomistic Crystals

The approach outlined above can be translated into a simple computer algorithm to detect and find all dislocations in an atomistic crystal (Stukowski 2014). We
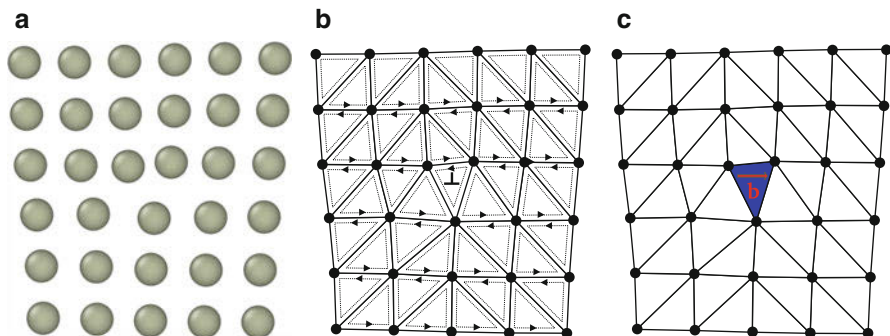
**Fig. 4** (**a**) Input atomic positions. (**b**) Tessellation of the input domain into triangular circuits using the Delaunay construction. (**c**) Burgers circuit test, performed on each Delaunay triangle, reveals exactly one cell that contains the dislocation

assume that the crystal to be analyzed is specified as a set of atomic coordinates $\{\mathbf{x}_i\}$ as depicted in Fig. 4a. The Delaunay construction is used to tessellate the crystal domain into a set of triangles (Fig. 4b). The elements of the triangulation are space-filling and non-overlapping, and we can regard them as small, elementary Burgers circuits, which will allow us to find and locate all dislocations contained in the atomistic crystal.

Since every edge of a triangle *abc* of the Delaunay tessellation represents an atom-to-atom step, we can apply the discrete version of the Burgers circuit method to calculate the per-triangle closure failure $\mathbf{b}_{abc} = \Delta\mathbf{x}'_{ab} + \Delta\mathbf{x}'_{bc} + \Delta\mathbf{x}'_{ca}$ after mapping each edge vector $\Delta\mathbf{x}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ connecting two successive atoms $i$ and $j$ to its corresponding ideal vector $\Delta\mathbf{x}'_{ij}$ in a perfect reference crystal lattice. Triangle circuits with closure failure $\mathbf{b}_{abc} \neq 0$ are marked as containing a dislocation, as shown in Fig. 4c.

There are different ways to accomplish the mapping of interatomic vectors from the dislocated crystal to the virtual reference lattice, $\Delta\mathbf{x}_{ij} \rightarrow \Delta\mathbf{x}'_{ij}$. In cases where the orientation of the dislocated crystal in the simulation coordinate system is known a priori, we can simply pick the vector $\Delta\mathbf{x}'_{ij}$ from a prescribed set of ideal lattice vectors taking the one that is closest to the elastically distorted vector $\Delta\mathbf{x}_{ij}$ (Stukowski 2014). In more general situations, a structure identification method such as common neighbor analysis (CNA) (Honeycutt and Andersen 1987; Faken and Jonsson 1994) or polyhedral template matching (PTM) (Larsen et al. 2016) must be used to first determine the local lattice orientation and then map atomic neighbor vectors to corresponding ideal lattice directions.

So far, we have considered only two-dimensional crystals where dislocations are point-like object in the plane. How does this approach extend to three-dimensional crystals containing linear dislocations? Here, the Delaunay tessellation of the atomistic model consists of tetrahedral cells, each being bordered by four triangular facets (Fig. 5). In the three-dimensional version of the algorithm, the closure failure $\mathbf{b}_{abc}$ must be computed for every triangular facet of the tetrahedral Delaunay cells.
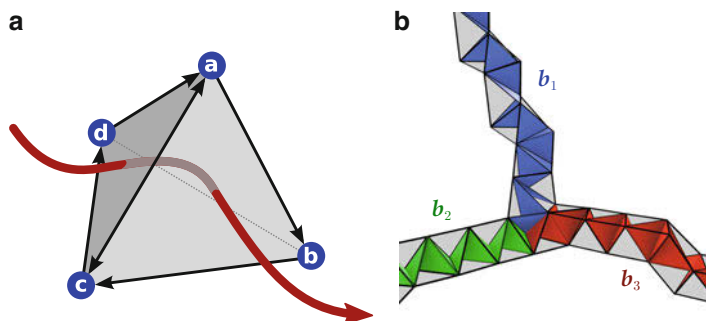
**Fig. 5** (**a**) A tetrahedral cell of the three-dimensional Delaunay tessellation, which is spanned by four vertex atoms. A dislocation line enters and exits through the triangular facets of the cell. The algorithm described in the text identifies such facets using the Burgers circuit test. (**b**) Linear defects with dislocation character lead to "chains" of dislocated Delaunay cells, which may form junctions in three-dimensions as exemplarily shown here

If $\mathbf{b}_{abc} \neq 0$, a facet is marked as being intersected by a dislocation line. Since dislocations cannot end within an otherwise perfect crystal, because of the Burgers vector conservation law, a line entering a Delaunay cell through one of its triangular facets must exit the cell again through one of its other three facets. Accordingly, the dislocation can be viewed as a line piercing through a sequence of triangular facets and tetrahedral cells as illustrated by Fig. 5b.

## 4    Dislocation Extraction Algorithm (DXA)

So far we have deliberately ignored several important aspects that can play a role in more general situations. First and foremost, crystal dislocations have a finite core size (Fig. 6a). That means they are not mathematically thin, one-dimensional objects but rather tubelike objects spread over a certain space region. The core region typically extends over more than one interatomic spacing and is thus covering more than one Delaunay triangle element. In order to capture such dislocations, larger Burgers circuits are necessary to fully enclose the core (which is represented by a connected set of Delaunay elements that have been marked as "bad"; see Fig. 6b).

Secondly, dislocations may dissociate into partial dislocations. If we want to identify partial dislocations individually, e.g., Shockley partials in fcc crystals, using the Burgers circuit procedure, we have to take special provisions, as the circuit enclosing the dislocation necessarily passes through the adjacent stacking fault defect (Fig. 6c). Only if we map the atomic step leading through that stacking fault plane to the correct fractional lattice vector, we will obtain the right (fractional) Burgers vector of the partial dislocation.

Finally, crystals often contain other defects in addition to dislocations. A general dislocation identification algorithm must therefore be able to deal with non-dislocation irregularities such as free surfaces, point defects, grain boundaries, other
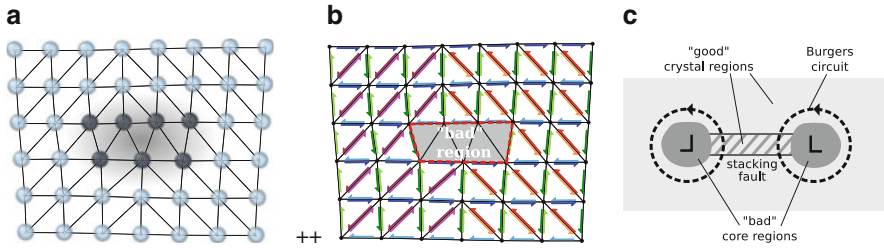
**Fig. 6** (**a**) Dislocation with an extended core. Atoms that are part of the core (darker color) can be identified using a structural characterization technique such as common neighbor analysis. (**b**) Bad tessellation elements, for which no unambiguous mapping to the perfect reference lattice is possible, have been marked with a gray color. (**c**) Schematic depiction of a dissociated dislocation. Identification of the two partial dislocations requires Burgers circuits passing through the stacking fault

types of interfaces, and even disclinations in a robust way. And, as mentioned before, the crystal orientation and crystal structure may not be known in advance and can vary across space and time. The algorithm needs to adapt to these situations appropriately.

In order to address these challenges, a computer algorithm named the *dislocation extraction algorithm (DXA)* (Stukowski et al. 2012; Stukowski and Albe 2010) has been devised on the basis of the fundamental ideas described in the preceding sections. The DXA is capable of building a discretized line representation of all dislocations contained in a given atomistic crystal model (Fig. 1). The generated representation of dislocation lines found in the crystal is very similar to those employed by dislocation dynamics simulation models. The DXA is available as part of the OVITO (Stukowski 2010) data analysis and visualization software for atomistic simulations.

The DXA proceeds in several steps, starting with the atomic input coordinates, to arrive at the final line representation of the dislocations. Here is a synopsis of these processing steps, which are described in more detail in Stukowski et al. (2012):

1. The three-dimensional Delaunay tessellation is computed.
2. Atoms in the input crystal are identified that form a perfect crystal lattice. For this, the common neighbor analysis (CNA) method is used, which allows to identify common lattice types such as fcc, bcc, hcp, and diamond. The information is also used to determine local lattice orientations and map atom-to-atom vectors in the Delaunay tessellation to the ad hoc reference lattice.
3. Elements in the Delaunay tessellation are flagged as "bad" crystal regions if they contain disordered atomic arrangements. This includes the dislocation cores, where the atomic structure deviates considerably from one of the perfect crystals, but also other types of defects (Fig. 6b).
4. The separating surface between the "good" and the "bad" crystal regions in the Delaunay tessellation, the so-called interface mesh, is generated (see Fig. 7).
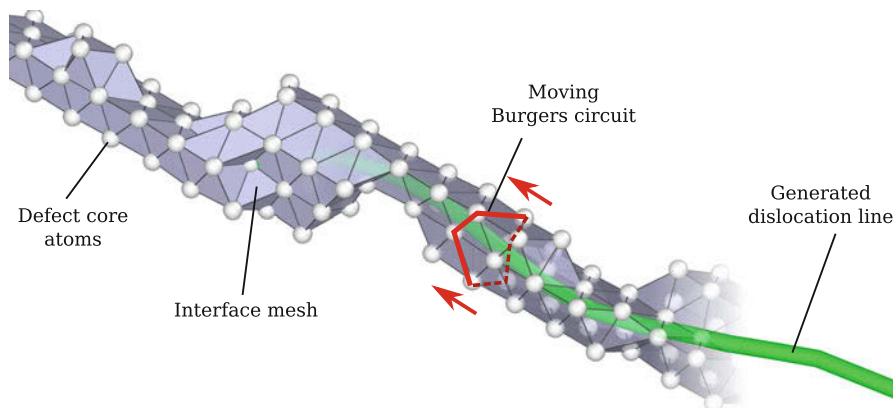
**Fig. 7** Illustration of the interface mesh constructed by the DXA to enclose all defect atoms of a dislocation core. The algorithm uses an "elastic" Burgers circuit (red) that is moving on the interface mesh to sweep the dislocation line. While this circuit is being advanced in a stepwise fashion, triangle by triangle, a continuous line representation of the dislocation defect is produced (green)

5. The algorithm then generates a large number of trial circuits on the interface mesh until it encounters a first circuit that fully encloses a dislocation. This is detected by computing the Burgers sum (Eq. 1). The maximum size of the trial circuits is bounded by a user parameter controlling how wide dislocation cores may be for the algorithm to detect them.
6. The first circuit is subsequently used to discover the rest of the current dislocation line. This happens by advancing the circuit on the interface mesh and sweeping along the dislocation line as depicted in Fig. 7.
7. During this sweeping phase, a one-dimensional line representation of the dislocation is generated by computing the new center of mass of the circuit every time it is advanced along the boundary of the dislocation core. Here, a circuit can be pictured as a rubber band tightly wrapped around the dislocation's core. As the circuit moves along the dislocation segment, it may need to locally expand to sweep over wider sections of the core, e.g., kinks or jogs. To prevent the circuit from sweeping past dislocation junctions or interfaces, again a limit is imposed on the circuit length.
8. As a last step, a post-processing of the discretized dislocation lines is performed to reduce the number of sampling points.

The sweeping of dislocation lines, performed in steps 6 and 7 of the algorithm, in fact happens simultaneously on all segments of a dislocation network as depicted in Fig. 8. The initial seed circuits, constructed at the slimmest spots of the dislocation segments, split into pairs of circuits, each sweeping along the cores' surfaces in opposite directions. During this sweeping process, the upper limit for each circuit's maximum length is continuously raised, letting the circuits approach closer and closer to the dislocation junctions, which typically exhibit a wider cross section than
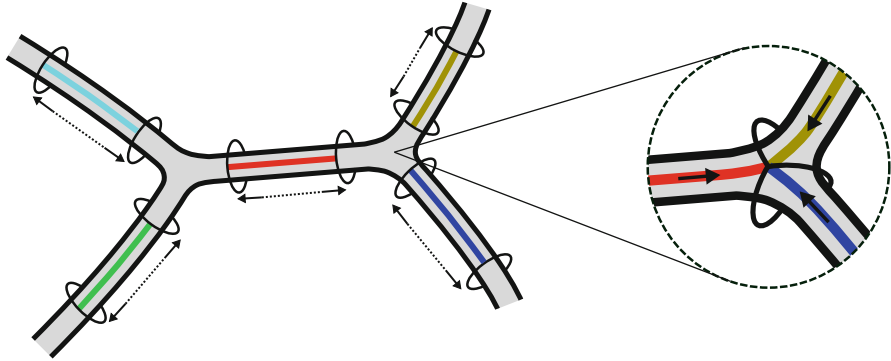
**Fig. 8** Schematic depiction of the DXA line tracing process for a network of dislocations. All dislocation arms are simultaneously swept by pairs of Burgers circuits advancing on the core's surface in opposite directions (cf. Fig. 7). At junctions, the inbound circuits from different arms meet, and the algorithm outputs a nodal point to connect all lines traced by these circuits
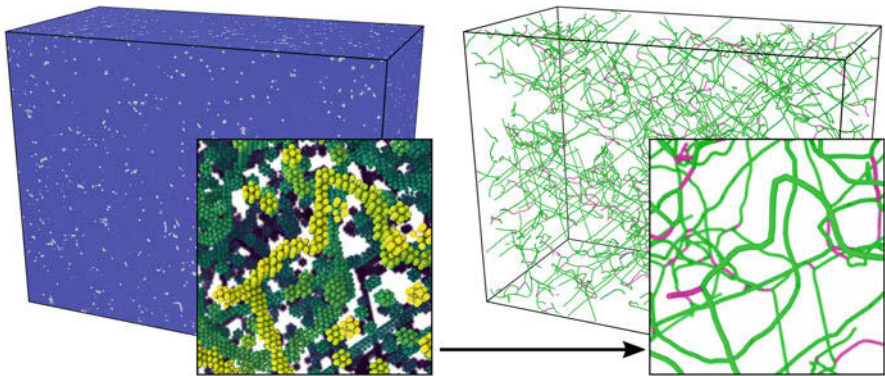


**Fig. 9** Left: Molecular dynamics simulation of dislocation-based single crystal plasticity (Zepeda-Ruiz et al. 2017). The piece of tantalum crystal (33 million atoms) is being deformed under uniaxial compression. Common neighbor analysis filtering (Stukowski 2012) of the atomic coordinates reveals a high density of defects in the bcc lattice (inset). Right: After processing with the DXA, non-dislocation defects such as vacancies have been filtered out. The resulting line representation allows measuring dislocation densities and studying dislocation processes in great detail. Green and magenta lines represent $\frac{1}{2} \langle 111 \rangle$ and $\langle 100 \rangle$ dislocations, respectively

the dislocation arms. At some point, the converging circuits all meet in a junction, and the algorithm links up their corresponding line ends at a nodal point.

## 5    Use Cases of the DXA

The DXA can serve as a measurement tool to quantify the density of dislocations in molecular dynamics simulations, thanks to the conversion of the identified defects to a mathematical line representation (Fig. 9). The average dislocation density in a
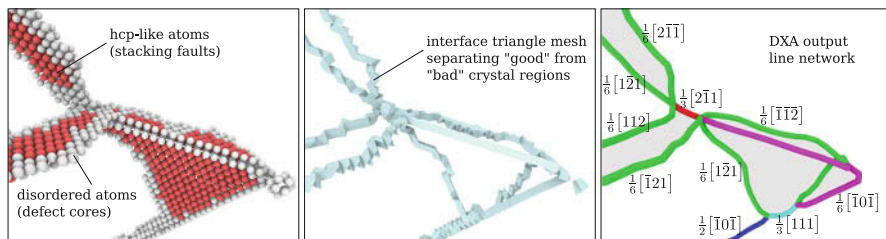
**Fig. 10** Analysis of a complex dislocation junction in fcc Al that has formed through a reaction of two $\frac{1}{2}\langle 110\rangle$ dislocation loops on intersecting glide planes. Left: Atomic input configuration with applied CNA filtering to highlight stacking fault atoms (red) and defect core atoms (gray). Center: Visualization of the *interface mesh*, the intermediate structure constructed by the DXA to enclose the defect cores. Right: Output dislocation lines and Burgers vector labeling generated by the DXA

crystal, $\rho$, is simply computed from the integral line length divided by the simulation box volume (see, e.g., Zepeda-Ruiz et al. 2017). The generated description of the dislocation geometry may also be used, for instance, to measure the size of the plastic zone in nanoindentation simulations (Gao et al. 2015; Voyiadjis and Yaghoobi 2015; Remington et al. 2014; Yaghoobi and Voyiadjis 2016; Alhafez et al. 2016; Alabd Alhafez et al. 2017) or to obtain local dislocation densities. Since the DXA does not only yield the shape but also the Burgers vector of each dislocation segment, the total (or statistical) density as well as the geometrically necessary density of dislocations (GND) are accessible via the DXA.

Another typical use case of the DXA is detailed analyses of dislocation reactions in molecular dynamics and statics simulations (e.g., Zhang et al. 2017). These reactions include the formation and nucleation of new dislocations at other defects, e.g., free surfaces (Trushin et al. 2016), grain boundaries (Stukowski et al. 2010), crack tips (Vatne et al. 2013), or pores (Ruestes et al. 2014) in a material. Here, the DXA can generate a precise labeling of dislocations forming complex network configurations as demonstrated in Fig. 10.

## 6    Current Limitations of the DXA

While the DXA represents a great improvement over conventional, atom-based analysis techniques such as the CNA, it still has certain limitations that one should be aware of when using this analysis tool. This section provides a roundup of key issues that need to be taken into account while working with the DXA and its output. Note, however, that the DXA is the subject of ongoing research seeking to improve the algorithm and overcome some of the issues mentioned here.

**Accuracy and ambiguity of dislocation representations** In general, given an atomistic crystal configuration, the representation of the contained defects in terms of a set of discrete dislocation lines is not uniquely defined. For instance,

dislocations in fcc crystals typically dissociate into pairs of partial dislocations (see upper left corner in Fig. 10). If the separation distance between such two partials becomes very small (on the order of 1–2 interatomic spacings), it is up to the algorithm to decide whether this particular atomistic configuration is better represented by a single dislocation line or two partials. The algorithm's choice is based on its particular notion of the dislocation core (in DXA terminology termed the "bad" crystal region, following Frank 1951). If the two dislocation core regions overlap, the algorithm is only able to construct a single Burgers circuit enclosing both defects. As a result, the two dislocations are fused into a single discrete dislocation line in the DXA's output representation. On the other hand, if the two cores are separated by just some "good" crystal region in between, then the DXA generates Burgers circuits around each of the individual partials, and two separate lines will be generated in the output. Thus, some "good" crystal atoms are required in between the two dislocations to separate them, cf. Figs. 6c and 10. A second type of ambiguity arises for very short dislocation segments and at dislocation junctions. Figure 11 depicts a detail of a dislocation network where four arms merge into a junction (a "4-junction"). However, they do not meet exactly in one point, and due to this slight dissociation of the junction, there is freedom of whether to describe this configuration as two separate 3-junctions instead, which are connected by an additional short dislocation segment. Which topology the DXA prefers depends on minutiae of the core morphology at this junction. As the four inbound Burgers circuits approach the junction, the upper limit on the circuits' lengths is continuously raised, letting the circuits stretch and advance further into the junction step by step. Simultaneously, the algorithm tries to generate an additional seed circuit around the core of the dissociated junction, i.e., in the inner area that has not been swept by the existing circuits yet. If the algorithm succeeds in spawning another Burgers circuit for the connecting segment before the existing circuits have met in the junction, then a topology with two 3-junctions results. Otherwise, the algorithm yields a single 4-junction. Effectively, the outcome is determined by the ratio of the core diameter of the dissociation segment and its length.

**Supported crystal structures** The DXA relies on an ad hoc mapping being established between the dislocated crystal and a corresponding ideal reference crystal lattice. This mapping is accomplished by means of an atomic structure identification method (Stukowski 2012), which maps the nearest neighbor vectors of each identified input atom to corresponding ideal lattice directions. The current implementation of the DXA uses the common neighbor analysis (CNA) method (Honeycutt and Andersen 1987) as a subroutine for this step, and it is thus currently limited to bcc, fcc, and hcp crystals and, thanks to a recent extension of the CNA method (Maras et al. 2016), also cubic and hexagonal diamond structures. Since the chemical types of atoms are not relevant to the DXA itself, crystalline compounds with a sub-lattice matching one of these supported structure types can also be processed by feeding only atoms from a sub-lattice to the algorithm. It is expected that future implementations of the DXA will employ new structure identification techniques other than the CNA in order to support a wider range of crystal structures.
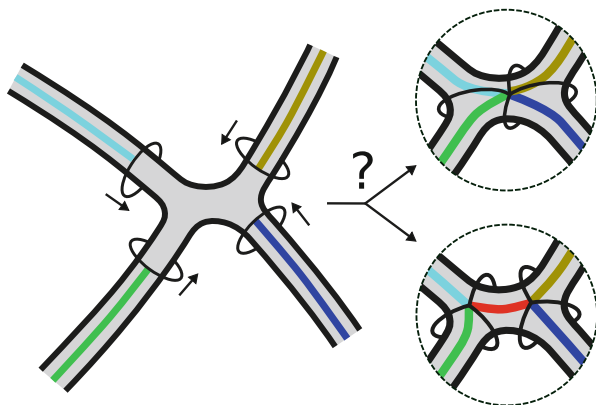
**Fig. 11** At slightly dissociated dislocation junctions like the one shown here, an ambiguity arises of whether to connect all arms at a single node or instead create two nodes that are connected by a short extra segment

In principle, any computational method that establishes a local mapping between atomic neighbor vectors and the ideal reference lattice can serve as foundation for the dislocation detection part of the DXA.

**Glide plane identification** The DXA yields the geometric shape of a dislocation as well as its Burgers vector. This information alone, however, is not sufficient to identify the glide plane the dislocation is moving on. Thus, the DXA falls short of answering the question of active slip systems in a deforming crystal. In certain cases, however, one can use heuristic criteria to guess the glide planes of dislocations. For Shockley partials in fcc crystals, a dislocation's Burgers vector uniquely determines its glide plane, and no other information is needed. In other cases, if a dislocation has an edge component (is not pure screw), its glide plane can be determined from the Burgers vector and the line direction. In general, however, it is important to recognize that the glide plane of a dislocation is a dynamic property and requires the analysis of the dislocation's path it takes through the crystal, which is beyond the DXA's capabilities (see next item).

**Tracking dislocations through time and space** It is important to note that the DXA operates on instantaneous snapshots of an atomistic crystal and builds a line model of the dislocations at certain simulation times. In other words, it is a static analysis method, not a kinematic one. Since dislocations are not physical objects (see our introductory discussion), they do not possess unique identities that would allow to track them over time. This makes it difficult to automatically correlate successive snapshots of the evolving dislocation configuration as dislocations can move arbitrary distances between MD simulation snapshots, undergo reactions, and appear newly via nucleation and disappear via absorption or annihilation processes. In particular, the DXA cannot directly deliver dislocation velocity

information, because tracking of dislocations would require additional heuristics to link dislocations in successive DXA snapshots.

# References

Alabd Alhafez I, Ruestes CJ, Urbassek HM (2017) Size of the plastic zone produced by nanoscratching. Tribol Lett 66(1):20. https://doi.org/10.1007/s11249-017-0967-9

Alhafez IA, Ruestes CJ, Gao Y, Urbassek HM (2016) Nanoindentation of hcp metals: a comparative simulation study of the evolution of dislocation networks. Nanotechnology 27(4):045706. http://stacks.iop.org/0957-4484/27/i=4/a=045706

Bulatov VV, Cai W (2006) Computer simulations of dislocations. Oxford University Press, Oxford/New York

Faken D, Jonsson H (1994) Systematic analysis of local atomic structure combined with 3D computer graphics. Comput Mater Sci 2(2):279–286

Frank FC (1951) LXXXIII. Crystal dislocations – elementary concepts and definitions. Philos Mag Ser 7 42(331):809–819

Gao Y, Ruestes CJ, Tramontina DR, Urbassek HM (2015) Comparative simulation study of the structure of the plastic zone produced by nanoindentation. J Mech Phys Solids 75(Suppl C):58–75. https://doi.org/10.1016/j.jmps.2014.11.005, http://www.sciencedirect.com/science/article/pii/S00225096140%0221X

Honeycutt JD, Andersen HC (1987) Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. J Phys Chem 91(19):4950–4963

Larsen PM, Schmidt S, Schiøtz J (2016) Robust structural identification via polyhedral template matching. Model Simul Mater Sci Eng 24(5):055007. http://stacks.iop.org/0965-0393/24/i=5/a=055007

Maras E, Trushin O, Stukowski A, Ala-Nissila T, Jónsson H (2016) Global transition path search for dislocation formation in Ge on Si(001). Comput Phys Commun 205(Suppl C):13–21. https://doi.org/10.1016/j.cpc.2016.04.001, http://www.sciencedirect.com/science/article/pii/S00104655163%00893

Remington T, Ruestes C, Bringa E, Remington B, Lu C, Kad B, Meyers M (2014) Plastic deformation in nanoindentation of tantalum: a new mechanism for prismatic loop formation. Acta Mater 78(Suppl C):378–393. https://doi.org/10.1016/j.actamat.2014.06.058, http://www.sciencedirect.com/science/article/pii/S13596454140%04881

Ruestes C, Bringa E, Stukowski A, Nieva JR, Tang Y, Meyers M (2014) Plastic deformation of a porous bcc metal containing nanometer sized voids. Comput Mater Sci 88(Suppl C):92–102. https://doi.org/10.1016/j.commatsci.2014.02.047, http://www.sciencedirect.com/science/article/pii/S09270256140%01517

Stukowski A (2010) Visualization and analysis of atomistic simulation data with OVITO – the open visualization tool. Model Simul Mater Sci Eng 18(1):015012. https://doi.org/10.1088/0965-0393/18/1/015012

Stukowski A (2012) Structure identification methods for atomistic simulations of crystalline materials. Model Simul Mater Sci Eng 20(4):045021. https://doi.org/10.1088/0965-0393/20/4/045021

Stukowski A (2014) A triangulation-based method to identify dislocations in atomistic models. J Mech Phys Solids 70:314–319. https://doi.org/10.1016/j.jmps.2014.06.009, http://www.sciencedirect.com/science/article/pii/S00225096140%01331

Stukowski A, Albe K (2010) Extracting dislocations and non-dislocation crystal defects from atomistic simulation data. Model Simul Mater Sci Eng 18(8):085001. https://doi.org/10.1088/0965-0393/18/8/085001

Stukowski A, Albe K, Farkas D (2010) Nanotwinned fcc metals: strengthening versus softening mechanisms. Phys Rev B 82:224103. https://doi.org/10.1103/PhysRevB.82.224103

Stukowski A, Bulatov V, Arsenlis A (2012) Automated identification and indexing of dislocations in crystal interfaces. Model Simul Mater Sci 20:085007

Trushin O, Maras E, Stukowski A, Granato E, Ying SC, Jónsson H, Ala-Nissila T (2016) Minimum energy path for the nucleation of misfit dislocations in Ge/Si(0 0 1) heteroepitaxy. Model Simul Mater Sci Eng 24(3):035007. http://stacks.iop.org/0965-0393/24/i=3/a=035007

Vatne IR, Stukowski A, Thaulow C, østby E, Marian J (2013) Three-dimensional crack initiation mechanisms in bcc-Fe under loading modes I, II and III. Mater Sci Eng A 560(Suppl C):306–314. https://doi.org/10.1016/j.msea.2012.09.071, http://www.sciencedirect.com/science/article/pii/S09215093120%13858

Voyiadjis GZ, Yaghoobi M (2015) Large scale atomistic simulation of size effects during nanoindentation: dislocation length and hardness. Mater Sci Eng A 634(Suppl C):20–31. https://doi.org/10.1016/j.msea.2015.03.024, http://www.sciencedirect.com/science/article/pii/S09215093150%02579

Yaghoobi M, Voyiadjis GZ (2016) Atomistic simulation of size effects in single-crystalline metals of confined volumes during nanoindentation. Comput Mater Sci 111(Suppl C):64–73. https://doi.org/10.1016/j.commatsci.2015.09.004, http://www.sciencedirect.com/science/article/pii/S09270256150%0573X

Zepeda-Ruiz LA, Stukowski A, Oppelstrup T, Bulatov VV (2017) Probing the limits of metal plasticity with molecular dynamics simulations. Nature 550:492

Zhang L, Lu C, Tieu K, Su L, Zhao X, Pei L (2017) Stacking fault tetrahedron induced plasticity in copper single crystal. Mater Sci Eng A 680(Suppl C):27–38. https://doi.org/10.1016/j.msea.2016.10.034, http://www.sciencedirect.com/science/article/pii/S09215093163%12424

# Line Dislocation Dynamics Simulations with Complex Physics

# 73

R. B. Sills and S. Aubry

## Contents

## Abstract

Discrete dislocation dynamics (DDD) simulations provide a technique for examining the effects of fundamental dislocation physics on the plastic response of crystalline solids. Many DDD simulations focus on relatively "simple" materials and loading conditions, such as glide-motion-dominated plasticity of pure cubic crystals. The goal of this chapter is to provide an overview of the more "complex" physical aspects of dislocation-mediated plasticity in the context of DDD. We

R. B. Sills (✉)
Sandia National Laboratories, Livermore, CA, USA
e-mail: rbsills@sandia.gov

S. Aubry
Lawrence Livermore National Laboratory, Livermore, CA, USA
e-mail: sylvie.aubry@llnl.gov

consider both physics that are intrinsic to the crystal lattice (elastic anisotropy, nonlinear drag, and low crystallographic symmetry) and extrinsic physics that are due to defects other than dislocations (solutes, vacancies, precipitates, and grain boundaries). For each of these classes of physics, we first discuss the conditions under which they are relevant, followed by an examination of the fundamental ways in which the behaviors of dislocations are affected by the physics, and finally a presentation of the methods that have been developed for incorporating the physics in DDD. We end the chapter by discussing three example simulations where complex physics are consequential.

# 1    Introduction

Plastic, or permanent, deformation in crystalline materials is often due to the motion of defects called dislocations. Dislocations are line defects – taking the shape of lines as they weave their way through a crystal – and their motion causes the two halves of a crystal on either side of the dislocation to shift relative to each other, thereby permanently changing the crystal's shape. Dislocations are characterized by the direction and magnitude of the shift they induce, known as the Burgers vector, $b$. They have long-range stress fields that cause two dislocations to interact with each other at a distance. Dislocations can also interact over very short distances and combine together to form new dislocation lines. Depending on the material, load state, and environmental conditions, dislocation motion is usually anisotropic, with a few preferred crystallographic directions of motion. In order to understand dislocation-mediated plasticity, we need an approach that can combine all of these aspects of dislocation physics into a versatile computational framework; this is precisely what is accomplished with discrete dislocation dynamics (DDD).

Using DDD, we study plasticity in crystalline materials by simulating the motion of individual dislocation lines in space and time. Usually the length scale of these simulations is on the order of several microns at the so-called mesoscale. At this length scale, we treat crystalline defects such as dislocations, precipitates, or grain boundaries as specialized objects embedded in an elastic medium. Treating the defects explicitly in this manner allows for their connection to mechanical properties to be accurately captured. However, unlike with atomic-scale simulations where the properties of defects are inherited from the governing interatomic interactions, the properties of defects at the mesoscale must be directly incorporated into the model framework. For example, the types of interactions between dislocations and precipitates must be defined and explicitly incorporated into the code. The benefit of the mesoscale approach is efficiency; while mesoscale simulations are lower fidelity than atomic-scale simulations, they can access length and time scales more relevant to experiments and applications. Most atomistic simulations examine the behavior of just one or a few dislocations, whereas DDD simulations may consider many thousands of dislocation lines.

For pure face-centered cubic (FCC) and body-centered cubic (BCC) metals, the physical models and algorithms used in DDD simulations are fairly well

established. This includes computation of dislocation-dislocation interaction forces using isotropic linear elasticity (Sect. 2.1), linear mobility laws for determining dislocation velocities (Sect. 2.2), and rules for junction formation (Sect. 2.3). These methods are well developed and thoroughly documented in a number of review papers, chapters, and textbooks (Tang 2005; Van der Giessen and Needleman 2005; Zbib and Khraishi 2005; Bulatov and Cai 2006; Arsenlis et al. 2007; Sills et al. 2016). However, most engineering materials are not pure FCC and BCC metals; most have comparatively complex physics governing their plastic response, and it is the study of these materials that is the focus of this chapter.

The goal of this chapter is to summarize and present methods for incorporating complex physics into a three-dimensional DDD simulation. This will necessarily involve some discussion of dislocation physics and strengthening mechanisms; however for more thorough treatments, readers are encouraged to study the works of Hirth and Lothe (1992), Hull and Bacon (2011), Argon (2008), and Kubin (2013). Our primary aim is threefold: (1) inform the reader of physics relevant to a broad range of materials and applications, (2) discuss methods and algorithms for incorporating these physics into a DDD simulation, and (3) provide several examples where complex physics are important. A key point emphasized throughout is that the material system and application of interest will motivate the relevant physics. Since DDD is a mesoscopic method, the physics of interest must be explicitly added into a simulation, and this usually requires significant time and effort; it is important to scope out the problem appropriately from the get-go to avoid wasted time.

The balance of the chapter is organized as follows. We will first concisely summarize basic DDD methods – those applied to "simple" materials. We will then discuss the ways in which additional physics may be incorporated into a DDD simulation. The bulk of the chapter focuses on discussing the different types of "complex" physics, divided into two categories: intrinsic physics, meaning physics intrinsic to the underlying crystal lattice, and extrinsic physics or physics that are due to crystal defects other than dislocations. Finally, we provide a few examples before concluding the chapter.

## 2 Basic Physics

### 2.1 Driving Forces

In discrete dislocation dynamics, the dislocation structure can be discretized and represented by a network of dislocation segments connected at nodes. The dislocation line network evolves under the influence of driving forces acting on each segment. In an infinite medium, the origin of these driving forces is threefold:

1. The force due to the external applied stress $\sigma_{\text{ext}}$. This is given by the Peach-Koehler formula for the force per unit length on a dislocation with Burgers vector $\boldsymbol{b}$ and line direction $\boldsymbol{\xi}$:

$$F = (\sigma_{\text{ext}} \cdot b) \times \xi. \tag{1}$$

2. The elastic force due to the stress fields of the rest of the dislocation network. Computation of this force is the most expensive aspect of the DDD method. Arsenlis et al. (2007) showed that in isotropic elasticity the elastic force can be computed analytically.
3. The dislocation self-force acting on a segment. This is caused by the Peach-Koehler force from the segment's own stress field. This self-force is perpendicular to the segment and tends to rotate the segment away from high-energy orientations. When an orientation-independent core energy is added, the self-force acquires a longitudinal component, which tends to shorten the segment.

The expressions used for elastic force computation apply strictly to an infinite, homogeneous medium. When determining the interaction forces in heterogeneous materials, such as multi-materials, multiphase metals, or materials containing precipitates or solutes (see Sects. 3.2.1 and 3.2.3), additional driving forces are needed to account for the presence of interfaces or boundaries between the different material domains. These forces are known as *image forces* and are necessary to ensure that the boundary conditions between the different regions of the domain are properly enforced. Similarly, in domains containing free surfaces (e.g., half-spaces, thin films (Weinberger et al. 2009b), and cylinders (Weinberger and Cai 2007; Weinberger et al. 2009a), or for arbitrary geometries (Van der Giessen and Needleman 1995; Lubarda et al. 1993; Crone et al. 2014)), image forces are necessary to ensure that free surfaces are traction-free. Lastly, in anisotropic, polycrystalline materials (as most are, see Sects. 3.1.1 and 3.2.4), image forces arise to enforce compatibility between the grains. In all of these cases, a boundary value problem can be defined to determine the image forces and can be solved using a numerical technique (such as the finite element method).

## 2.2 Mobility Laws

In DDD, we typically assume overdamped dynamics such that the inertia of the dislocation lines can be neglected. This leads to equations of motion of the form

$$F_i^{\text{drive}} + F_i^{\text{drag}}(v_i) = 0, \tag{2}$$

where $F_i^{\text{drive}}$ is the total driving force acting on dislocation node/segment $i$, $F_i^{\text{drag}}$ is the total drag force resisting its motion, and $v_i$ is the velocity of the node/segment. Since the drag force depends on the velocity, Eq. (2) relates the driving force to the dislocation velocity. Many sources of drag may exist in a crystalline solid arising from the intrinsic properties of the crystal lattice (e.g., phonon dispersion) or from extrinsic features of the solid, such as solute atoms (see Sect. 3.2.1).

The response of dislocation lines to the driving force $\boldsymbol{F}_i^{\text{drive}}$ is defined by the dislocation mobility function that may depend on crystal type, local stress state, and temperature, as well as on the dislocation's geometry (e.g., Burgers vector, glide plane, segment orientation). Although in general dislocation mobility laws are nonlinear, simpler linear mobility laws have been proposed for BCC (Arsenlis et al. 2007) and FCC (Cai and Bulatov 2004) crystals.

Assuming that Eq. (2) can be solved explicitly for the velocity, we typically express the mobility law as

$$\boldsymbol{v}_i = \mathscr{M}_i(\boldsymbol{F}_i^{\text{drive}}), \tag{3}$$

where $\mathscr{M}_i$ is a generally nonlinear vector function. When the mobility function is linear, Eq. (3) simplifies to

$$\boldsymbol{v}_i = \boldsymbol{M}_i \cdot \boldsymbol{F}_i^{\text{drive}} \tag{4}$$

where $\boldsymbol{M}$ is a $3 \times 3$ matrix (its dependence on external conditions and the geometry of the system is omitted in Eq. (4)).

In principle, molecular dynamics (MD) simulations can supply mobility $\mathscr{M}$ functions for any point $\boldsymbol{x}$ on a dislocation line taking into account the glide and climb (refer to Sect. 3.2.2) of dislocations of arbitrary orientation (character). However, in practice, MD dislocation mobility data is available only for a limited set of dislocation characters and straining conditions. Commonly, mobility simulations are conducted only for pure edge and pure screw dislocations in a few specific slip planes (see, e.g., Groh et al. 2009 and Gilbert et al. 2011). Given mobility data obtained from either MD simulations or experiments, interpolation and extrapolation must be used to complete a dislocation mobility function for DDD simulations over the rest of orientation and velocity space.

Two families of drag functions – planar and pencil-glide mobility laws – are often used in DDD simulations depending on the crystal structure of the material, with the difference between them being how screw dislocations are treated. In planar mobility laws, the velocity of a screw dislocation is constrained to remain in its glide plane until a cross-slip operation occurs. This approach is suitable when cross-slip is a thermally activated, rare event, typically handled in DDD using a Monte Carlo implementation (Kubin et al. 1992). In pencil-glide mobility laws, screw dislocations are able to move freely in any plane that contains their dislocation line, and the notion of a glide plane becomes ambiguous. Two orthogonal glide directions (both orthogonal to the dislocation line) need to be defined to compute the screw mobility. For instance, in BCC crystals at elevated temperatures, dislocations are observed to follow a pencil-glide-type behavior, and the two glide directions are usually a nonlinear function of projection of the velocity along the {110} and the {112} family planes. In both the planar and pencil-glide approach, an edge mobility is defined for motion in the plane containing the Burgers vector and the dislocation line (the glide plane) and for motion out of this plane (climb).

## 2.3 Junctions

When two dislocation lines come into contact with each other, they are able to combine together and react to form a new dislocation called a junction dislocation. Not all dislocation intersections result in junction formation; the junction reaction only proceeds if the formation of the junction results in a reduction of the system's energy. In DDD, a number of algorithms have been developed to test for this condition (i.e., a junction formation operator) (Weygand et al. 2002; Schwarz 2003; Bulatov and Cai 2006). One example is the principle of maximum dissipation (Bulatov and Cai 2006; Arsenlis et al. 2007), which says that a junction should only be formed if doing so maximizes the dissipation rate, given by the sum of the dot products between the force and velocity vectors of each the dislocation segments. And finally, in materials with low crystallographic symmetry (e.g., hexagonal close-packed materials), junctions may form with "large" Burgers vectors, and these junctions may then subsequently "splinter" apart into a new pair of dislocation lines. We will discuss this mechanism in Sect. 3.1.3 and provide an example of junction splintering in Sect. 4.2.

## 3 Complex Physics

The incorporation of complex physics into DDD simulations typically manifests itself in two ways: (1) a change in the driving forces on the dislocation lines and/or (2) a change in the rules governing how dislocation lines can move through the simulation cell. Some physics require changes to both.

In the following we will go through examples of complex physics associated with dislocation-mediated plasticity. Our primary focus is on examples that have already been the subject of DDD studies. Each example begins with a *Background* section where the physics of interest are discussed in the greater context of crystal plasticity, followed by a *Nature of the Physics* section where the basic physics relevant to dislocation dynamics are presented, and an *Incorporation in DDD* section where specific algorithms and formulations are discussed.

### 3.1 Intrinsic Physics

#### 3.1.1 Anisotropic Elastic Media

**Background**
DDD simulations often assume isotropic elasticity to compute stresses at points in the simulation volume and forces between dislocations. Strictly speaking, however, essentially all crystalline materials exhibit some degree of elastic anisotropy. The strength of the deviation from isotropy – typically expressed in terms of the anisotropy ratio, with a value of 1 indicating perfect isotropy – is dictated by the

crystal structure and the nature of the interactions between the atoms. Under the normal range of temperature and pressure, many materials have an anisotropy ratio that is not too far from 1. In cubic crystals, the anisotropy ratio is often defined as $A = \frac{2C_{44}}{C_{11}-C_{12}}$ where $C_{11}$, $C_{12}$, and $C_{44}$ are the cubic elastic constants. For some materials, using isotropic theory might be appropriate (Rhee et al. 2001); however this assumption is not always valid. For example, aluminum is very nearly isotropic with $A = 1.21$ at room temperature (Hirth and Lothe 1992), whereas at elevated temperatures around 1200 K, $\alpha$-Fe is highly anisotropic with $A = 7.4$ (Fitzgerald et al. 2012).

### Nature of the Physics

Two features of dislocation physics can be strongly impacted by elastic anisotropy: dislocation line tension and elastic interactions between dislocations. A dislocation's line tension controls how it bows out in response to an applied stress and is dictated by the orientation dependence of the elastic energy (screw vs. edge). Strong anisotropy can cause large elastic energy differences between edge and screw dislocations, causing the line tension to vary dramatically with character angle. For example, dislocation loops in $\alpha$-Fe exhibit sharp corners, a feature that cannot be captured using isotropic elasticity theory (Aubry et al. 2011). The elastic interactions between two dislocations, which dictate the strength of attraction or repulsion between them, are also affected by elastic anisotropy. One example is junction formation; dislocation junctions only form if the two lines are attracted to each other (see Sect. 2.3).

### Incorporation in DDD

In dislocation dynamics simulations, the stress at a point due to a dislocation segment and the interaction force between a pair of segments can be defined using linear elasticity. Its definition is based on the derivatives of the Green's function. While the Green's function is known analytically for isotropic media, no closed-form expression exists for anisotropic media. Hence, analytical stress and force expressions cannot be derived for anisotropic media, and instead they must be evaluated numerically for a given dislocation configuration.

In Rhee et al. (2001) and Yin et al. (2010), several approaches were explored for simulating anisotropic elastic media. In Yin et al. (2010), the authors show that the Willis-Steeds-Lothe expression (Hirth and Lothe 1992) is more computationally efficient than Brown's formula (Hirth and Lothe 1992) and that the matrix formalism is only slightly faster than the integral formalism. However, they also found that computing interaction forces in anisotropic elasticity was 220 times more computationally expensive than in isotropic elasticity. This fact motivated the development of a different approach by Aubry and Arsenlis (2013), where the spherical harmonics expansion of the derivative of the Green's function is derived. The advantage of writing the stress and the force expressions in terms of an expansion in spherical harmonics is that some integration can be carried out analytically, leading to improved efficiency. Within this approach, the accuracy and the computational cost of the calculations can be controlled by selecting the order of the expansion.

The cost of this method depends on the anisotropy ratio. For a small anisotropy ratio (less than 2, as with aluminum), the computational cost only increases by about 50% relative to isotropic elasticity. The higher the anisotropy ratio, the more expensive the method becomes. For instance, for $\alpha$-Fe at high temperature with an anisotropy ratio of 7.4, anisotropic simulations are about 60 times more expensive.

### 3.1.2 Nonlinear Lattice Drag

**Background**

Most of the time in DDD simulations, the velocity of a dislocation segment is assumed to be linearly related to the total driving force, i.e., the mobility law is linear as in Eq. (4). For many materials and many states of stress, this is a good approximation. However, under some conditions, nonlinear drag effects may become important. For example, it has been shown using MD simulations that in pure FCC (Olmsted et al. 2005; Marian and Caro 2006) and BCC (Barton et al. 2011) metals the force-velocity relation can be highly nonlinear at elevated stresses ($\gtrsim 100$ MPa). The physics governing these nonlinearities is complex and not entirely understood. Nonlinear drag effects can also be relevant when accounting for extrinsic sources of drag (like solute drag, see Sect. 3.2.1).

**Nature of the Physics**

A schematic showing different types of nonlinear mobility laws is presented in Fig. 1. In pure crystals, nonlinear drag is often due to non-phononic dissipative effects, such as radiative dissipation (Marian and Caro 2006). Oftentimes these non-phononic effects result in plateaus in the force-velocity relation, i.e., regions where the velocity becomes nearly independent of the force (Olmsted et al. 2005;



**Fig. 1** Schematic showing nonlinear mobility laws representing different drag mechanisms including linear drag, nonlinear lattice drag with velocity plateaus, solute friction (see Sect. 3.2.1), and kink-pair nucleation-limited motion
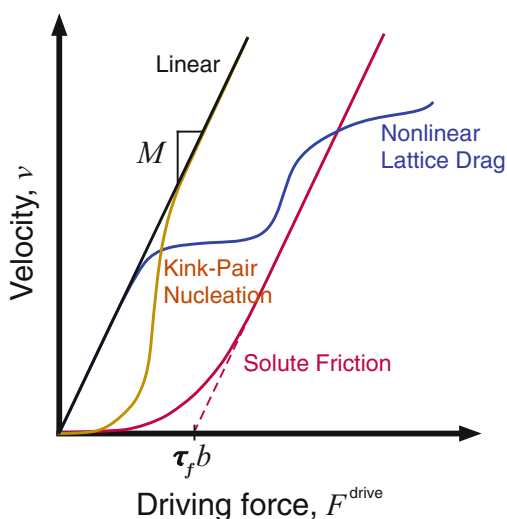
Marian and Caro 2006). Also, when the primary mechanism of dislocation motion is kink-pair nucleation and migration, as with BCC metals at low temperatures, a nonlinear mobility can result (Gilbert et al. 2011).

### Incorporation in DDD

Some closed-form nonlinear mobility laws have been implemented in DDD simulations (Po et al. 2016). However, because of the fact that we often determine nonlinear dislocation mobility laws using numerical data from MD, we rarely have a closed-form mobility law which can be explicitly solved for the velocity. This means that there are (at least) two issues that make usage of nonlinear mobility laws difficult: (1) interpolation of numerical MD data and (2) solution of an implicit, nonlinear system of equations. For example, cubic splines may be used to provide a continuous mobility function and Newton's method applied to iteratively solve the resulting nonlinear system. While neither of these issues is insurmountable, together they can make large-scale DDD simulations difficult. We recommend that nonlinear mobility laws are employed only if absolutely necessary.

### 3.1.3 Low-Symmetry Materials

### Background

The DDD method has predominantly been developed to evaluate the plastic behavior of cubic metals (e.g., FCC and BCC). These materials exhibit a high degree of crystalline symmetry, so that only a single class of Burgers vector is active ($\frac{1}{2}\langle 1\,1\,0 \rangle$ for FCC and $\frac{1}{2}\langle 1\,1\,1 \rangle$ for BCC). In order to consider crystals with relatively lower symmetry, such as the hexagonal close-packed (HCP), rhombohedral, or tetragonal crystal structures, additional physics are necessary to address the fact that more than one class of Burgers vector may be active.

### Nature of the Physics

In low-symmetry materials, some energetically favorable dislocation reactions lead to the formation of "composite" dislocations with Burgers vectors larger than the unit vectors of the crystal's lattice. In HCP crystals, such as magnesium or titanium, these large Burgers vectors have been associated with low ductility (Agnew et al. 2002; Wu and Curtin 2015). Formation of these composite Burgers vectors may not result in the greatest energy reduction for the system, however. It may be possible for the composite Burgers vectors to decompose into a new pair of dislocation lines that lead to a further energy reduction. This feature of low-symmetry materials motivates the need for new topological processes not necessary for cubic crystals.

### Incorporation in DDD

When the interaction force between the component dislocations that make up the composite dislocation is repulsive, the DDD algorithm should allow the splitting of the composite dislocation into its components. An additional topological operator, called "splintering," is needed to allow for the dissociation of composite dislocations

into two component dislocations when appropriate (Aubry et al. 2016). This operator behaves very similarly to the junction operator discussed in Sect. 2.3; an energy-based criterion is used to decide when and how it is appropriate to "splinter" a composite dislocation. An application of the splintering topological operation is shown in Sect. 4.2.

## 3.2 Extrinsic Physics

### 3.2.1 Solute Interactions

**Background**

We denote as "solutes" in a crystal the atoms that are different from those of the base lattice. They can exist as substitutional solutes occupying lattice sites of the base crystal or as interstitial solutes residing in lattice interstitial sites. The combined lattice-solute system is called a solid solution. Solid solutions arise in many settings. One example is alloying, the process of mixing multiple chemical species together in a crystalline solid. Most alloys are comprised of a base crystal with a few weight percent of an additional element mixed in, often in the form of a substitutional solid solution (Reed-Hill and Abbaschian 1992).

Forming a solid solution alters the behavior of the material in many ways. Our main focus here is on how solute atoms interact with dislocations to change their behavior, thereby altering the plastic response of a material – this phenomenon is often referred to as solid solution strengthening. Our key objective in studying solid solutions with DDD is to understand the major effects solute atoms have on the response of dislocation ensembles and the resulting impact on mechanical properties.

**Nature of the Physics**

Because solute atoms have a misfit in the parent lattice, they are commonly treated as misfitting Eshelby inclusions (Cai et al. 2014). As such, they have stress fields – both as a result of their misfit and the fact that they locally modify the modulus of the material – and can interact with dislocations through Peach-Koehler forces. Solutes are also able to interact with dislocations through interactions with the dislocation core (Medvedeva et al. 2005; Ventelon et al. 2015). Typically, the misfit effect is considered to be the dominant interaction mechanism (Argon 2008). As a consequence of this misfit, solutes tend to segregate around (predominantly edge) dislocations and form so-called Cottrell atmospheres.

The behavior of solid solution systems can be split into two categories based on the relative mobilities of the dislocations and solutes, as shown in Fig. 2. In the immobile solute case shown in Fig. 2a, the dislocations in a material are much more mobile than the solutes. In this case the solute population can often be treated as randomly distributed, and dislocation lines encounter each solute one by one as they move through the solid. Dislocation interactions with stationary solute atoms are usually considered to be thermally activated (Argon 2008). In contrast, in the
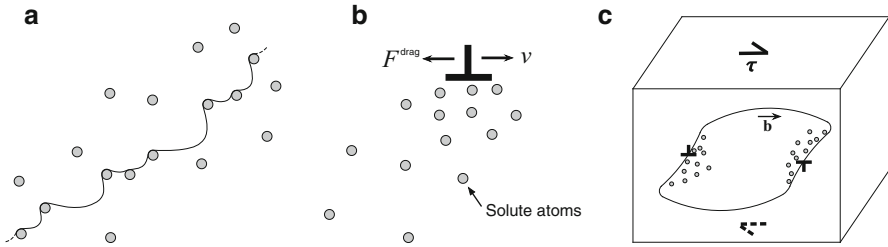
**Fig. 2** Schematics showing dislocation-solute interactions when (**a**) solutes are immobile and (**b**, **c**) solutes are mobile. (**a**) A dislocation line gliding through a field of stationary solutes. (**b**) A Cottrell atmosphere exerting a drag force on an edge dislocation gliding with velocity $v$. (**c**) A shear dislocation loop expanding under an applied stress with Cottrell atmospheres dragging the edge components

mobile solute case, the mobility of the solutes is greater than or comparable to that of the dislocations; Cottrell atmospheres are able to move with and exert drag forces on their dislocations. Figure 2 shows a schematic of a dragging Cottrell atmosphere with (b) a straight edge dislocation and (c) an expanding shear loop. The mobile solute regime is commonly referred to as the dynamic strain aging regime and is most relevant at elevated temperatures and with interstitial solutes, since solute diffusivities are largest under these conditions.

A number of solute effects on the properties of dislocations have been identified, including reduction in dislocation mobility (solute drag) (Cottrell and Jaswon 1949; Sills and Cai 2016; Varvenne et al. 2017), alteration of dislocation stress fields (Sofronis and Birnbaum 1994; Chateau et al. 2002; Cai et al. 2014), modification of the cross-slip rate (Wen et al. 2007; Nöhring and Curtin 2017), and reduction of line tension (Kirchheim 2007; Delafosse 2012; Sills and Cai 2018). Of these effects, the most significant is solute drag, which is our major focus here.

### Incorporation in DDD

One approach for incorporating solutes in a DDD simulation is to treat them as discrete solute objects. These solute objects can occupy predefined lattice positions set by the type of solid solution, with each having its own misfit-dependent stress field. They interact with dislocation segments through their stress fields, leading to solute drag forces. Mobile solutes can be accounted for using kinetic Monte Carlo methods, according to which solutes move by hopping between lattice sites with a rate set by the activation energy for diffusion of the solutes. This approach has been used to study solute impacts on junction formation and dissolution (Chen et al. 2008). The discrete solute approach is limited, however, by the large number of solutes contained within even a relatively small volume of material; the computational cost of the solutes can very quickly supersede that of the dislocations.

A less accurate but much more efficient approach is to incorporate solute drag into the mobility law. With immobile solutes, this is often accomplished by imposing a so-called friction stress, $\tau_f$, on the dislocations. In this model, a dislocation

segment is pinned by solutes and rendered immobile until the net driving force per unit length exceeds $\tau_f b$ (see Fig. 1). The magnitude of the friction stress is dependent upon the materials composing the solid solution, the solute concentration, the temperature, and the character of the dislocation line and can be determined with atomistic simulations (Olmsted et al. 2005). Friction stress models have been utilized by a number of researchers to incorporate solute drag effects (e.g., Monnet and Devincre 2006).

When the solutes are mobile, a friction stress model is inadequate. In this case, the effect of the dislocation velocity on the concentration field of solutes (Cottrell atmosphere) around the dislocation must be considered and the resulting drag forces computed. A nonlinear mobility law that incorporates these effects has been proposed on the basis of continuum theory calculations (Sills and Cai 2016); further research is necessary to better understand solute drag by mobile solutes.

### 3.2.2    Vacancy Interactions

**Background**
Vacancies are main crystal lattice sites where an atom is missing. Vacancies are always present in crystalline solids because they increase the entropy of the solid, with the vacancy concentration typically increasing with temperature. In fact, for a given material at a given temperature, there is a well-defined *equilibrium concentration* of vacancies. Many settings exist, however, where the vacancy concentration exceeds this equilibrium value. In these cases, the solid is said to be *supersaturated* with vacancies. Common examples include quenching from an elevated temperature and irradiation-damaged materials.

Although solute atoms and vacancies are both point defects, they interact with dislocations in very different ways. Vacancies primarily interact with dislocations by promoting *climb* motion, or motion out of the glide plane. In contrast to solutes, vacancies actually enhance dislocation mobility. In order for dislocation climb to have a significant impact on plasticity, the time scale of climb motion must be of the same order of magnitude as the time scale of interest. Hence, climb is typically a concern at elevated temperatures where vacancy diffusivities are large (and concentrations are higher) and/or when a material is under sustained load for an extended period of time (e.g., creep loading).

**Nature of the Physics**
As shown in Fig. 3a, in order for a non-screw dislocation line to move out of its glide plane (climb), vacancies or self-interstitials need to diffuse into or out of the dislocation core (depending on the climb direction). Mathematically, we define the *climb velocity* as the velocity component orthogonal to the glide plane, i.e., $v_{cl} = \mathbf{v} \cdot (\boldsymbol{\xi} \times \mathbf{b}/b_e)$, where $b_e$ is the magnitude of the edge component of the Burgers vector $\mathbf{b}$ and $\boldsymbol{\xi}$ is the dislocation line direction. The fundamental problem at hand is to determine the climb velocity for a given climb force, $f_{cl} = \mathbf{f} \cdot (\boldsymbol{\xi} \times \mathbf{b}/b_e)$, and vacancy concentration; this is akin to determining the "climb mobility law."

Vacancies can induce climb motion at a given point on a dislocation line by diffusing toward or away from the dislocation either through the bulk of the crystal *or* through the core of the climbing dislocation line. The latter is commonly referred to as *pipe diffusion* and is typically considered to be much more rapid than bulk diffusion (Hirth and Lothe 1992). In general, the climb velocity is related to the total vacancy flux into or away from the dislocation line per unit length, $I$, according to the relation (Hirth and Lothe 1992)

$$v_{cl} = \frac{I\Omega}{b_e},\qquad(5)$$

where $\Omega$ is the volume per atom. Determining this vacancy flux requires solving a coupled pipe-bulk diffusion problem. The problem is further complicated by the fact that dislocations act as sources and/or sinks for vacancies; one climbing dislocation line may act as a vacancy source for another climbing (sink) dislocation line as shown in Fig. 3b.

In order to determine the climb velocity, the climb force acting on the dislocation line must interact in some way with the vacancy concentration field, $c(\boldsymbol{x})$. This is usually accomplished by assuming that the vacancies and the dislocation line are in chemical equilibrium, leading to the relation $c_d = c_0 \exp\left[-f_{cl}\Omega/(b_e k_\mathrm{B} T)\right]$ (Hirth and Lothe 1992) where $c_0$ is the equilibrium vacancy concentration, $c_d$ is the concentration at the dislocation (on the surface of a tube of radius $r_d$ centered on the dislocation line), $T$ is the absolute temperature, and $k_\mathrm{B}$ is Boltzmann's constant. This relation represents the coupling between the dislocation microstructure and the vacancy concentration field. Finally, we note that a dislocation may have a nonzero climb velocity even when the stress state is zero if the vacancy concentration in the solid is not at equilibrium, as shown in Fig. 3c. In this case, we call the effective
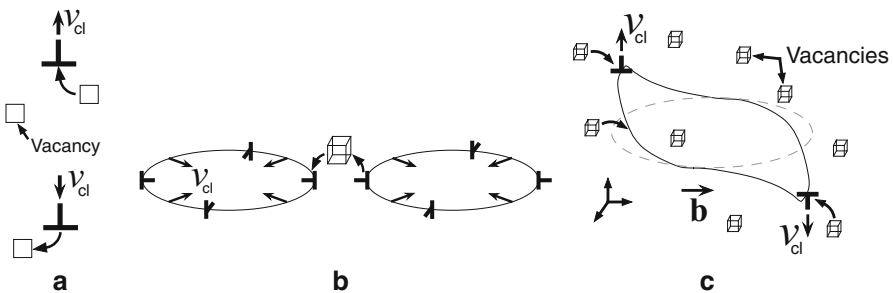


**Fig. 3** Schematics showing dislocation-vacancy interactions resulting in dislocations climbing with velocity $v_{cl}$. (**a**) Edge dislocations climbing with velocity "upward" by consuming vacancies and "downward" by emitting vacancies. (**b**) A pair of collapsing prismatic dislocation loops, with one loop emitting and the other absorbing vacancies. (**c**) A shear dislocation loop climbing out of its glide plane in the presence of a supersaturation of vacancies that produces osmotic forces

(thermodynamic) force exerted by the super- or subsaturated solid on the climbing dislocation line an *osmotic force*.

### Incorporation in DDD

While the physics governing vacancy-dislocation interactions are fairly complex, the overall process of computing the climb velocity is fairly straightforward: determine the vacancy flux $I$ at each point on the dislocation lines while enforcing the local equilibrium condition for $c_d$, and then compute the climb velocity with Eq. (5). All of the below approaches compute $I$ while assuming that the vacancy concentration field rapidly goes to steady state as the dislocation microstructure evolves.

The simplest approach is to neglect pipe diffusion and determine the vacancy flux by solving the steady-state bulk diffusion equation: $\nabla^2 c(\boldsymbol{x}) = 0$. In the earliest DDD implementations, the analytical solution for a straight, infinitely long edge dislocation in a cylinder of radius $r_\infty$ was employed (Mordehai et al. 2008). This solution is readily linearized, giving an expression of the form $v_{cl} = M_{cl}(f_{cl} + f_{os})$, where $M_{cl}(\theta)$ and $f_{os}(\theta)$ are the character-angle-dependent climb mobility and osmotic force, respectively (Bakó et al. 2011). In this approach, the dislocation lines are uncoupled from each other. In order to properly solve the diffusion equation, however, all of the sources and sinks (climbing dislocation lines) in the domain need to be coupled together. Obtaining such a solution is complex but can be accomplished readily using the Green's-function-based approach recently developed by Gu et al. (2015). In their method, the climb velocity for each dislocation segment can be determined by solving a coupled linear system.

Finally, none of these approaches account for pipe diffusion. Gao et al. (2011) implemented a pipe-diffusion-only climb model, finding that the climb velocity is proportional to the concentration gradient along the dislocation line. Recently, Niu et al. (2017) have derived a coupled pipe-bulk diffusion model by considering the micromechanics of vacancy diffusion, finding that the pipe contribution to the climb velocity is proportional to the second derivative of the core concentration field.

### 3.2.3    Precipitate Interactions

### Background

Precipitates are particles of a secondary phase embedded in a crystalline matrix. They come in many shapes and sizes, depending on how they are formed, the crystal structures involved, and the thermal and deformation history of the material. In many instances, they are intentionally introduced into a material in order to strengthen it, an approach that is called *precipitation strengthening*. Usually these precipitates are formed by first "solutionizing" an alloy by holding it at an elevated temperature and then subsequently "aging" it at a reduced temperature. During the aging process, solute atoms diffuse together and nucleate precipitates, which then continue to grow. Often, the properties and spatial distribution of the precipitates

change as they are aged. In other instances, precipitate formation is undesirable and can result in degraded mechanical properties. Finally, we can also think of voids, which can form during solidification, or gas-filled bubbles, which arise in radiation environments, as precipitates with a modulus of zero. Typically, voids and bubbles adversely affect mechanical properties. Whether favorable or not, the predominant way that precipitates alter a material's mechanical properties is through their interactions with dislocations. This makes DDD a valuable tool for studying precipitation-driven phenomena.

## Nature of the Physics

The mechanical properties of a precipitate-containing material are controlled by two things: the properties of the individual precipitates and their spatial distribution in the material. While individual precipitates are relatively well-understood, their distributions in space are difficult to characterize. Most often it is assumed that precipitates are randomly distributed; however some researchers have developed more sophisticated techniques for constructing realistic precipitate distributions (Mohles and Fruhstorfer 2002). It is also important to note that precipitates usually have a distribution of sizes as well. Some analytical models exist for incorporating a size distribution (Mohles 2004), or experimentally measured size distributions can be used (e.g., Boyd and Nicholson 1971).

Precipitates can interact with dislocations via both long-range and short-range mechanisms. We denote long-range interactions as those that do not require contact with a precipitate to act on the dislocation. Two types of long-range interactions exist. The first is due to the lattice mismatch of coherent precipitates, which gives rise to a lattice-mismatch stress field in both the precipitate and the matrix. The second type of long-range interaction exists if the elastic constants of the precipitate are different from those of the matrix; the resulting modulus mismatch gives rise to an additional stress field. Eshelby developed a mathematical procedure for computing the stress field of a precipitate with a lattice and/or modulus mismatch using elasticity theory (Eshelby 1961). Note that the stress field of a precipitate typically scales as $1/r^3$, in comparison to the (longer range) $1/r$ fields of the dislocation lines themselves.

Short-range interactions are present while a dislocation is in contact with and/or penetrating a precipitate. These interactions result from the resistance of a precipitate as a dislocation transmits into it. This transmission can only occur if the precipitate is coherent or semi-coherent. If a dislocation transmits all the way through a precipitate, we say that the precipitate was "cut" by the dislocation. We note that if a dislocation is unable to penetrate and cut a precipitate, it can still bypass the precipitate by bowing out and forming a dislocation (Orowan) loop around it. A number of mechanisms can give rise to contact interactions including stacking fault, modulus, or core energy differences between the precipitate and matrix, the formation of a step at the precipitate's surface, the disruption of superlattice ordering, and reactions with misfit dislocations (Argon 2008). Since these interactions occur over a very short length scale, they are dictated by nonlinear

interatomic interactions, making them difficult to characterize with continuum models. Atomistic simulations are currently the most accurate way to characterize the dislocation-precipitate cutting process (e.g., Singh and Warner 2013 or Monnet 2015).

### Incorporation in DDD

Precipitates are usually introduced into DDD as special objects that affect both the driving forces and mobility laws of the dislocation lines, with the specific details being dependent upon the properties of each precipitate. Typically the long-range interactions affect the driving forces, and the short-range interactions modify the mobility law.

In terms of long-range force calculation, the lattice-mismatch stress field is significantly easier to compute than the modulus-mismatch stress field. To compute the lattice-mismatch field of a single precipitate, all we need to know is the magnitude and character of the mismatch (expressed in terms of an eigenstrain tensor (Mura 1987)) and the size and shape of the precipitate. For many lattice mismatches and precipitate geometries, analytical solutions exist for the lattice-mismatch field. The total lattice-mismatch field is simply the superposition of the field from each precipitate in the problem domain. The resulting driving force can be computed using the Peach-Koehler formula.

When a modulus mismatch is present, the story is quite different. The modulus-mismatch field of a precipitate is dependent upon the type and magnitude of the modulus mismatch, the precipitate shape, *and* the stress field exerted on the precipitate by the applied stress and every other object in the domain, including all dislocation lines and all other precipitates. Calculating the modulus-mismatch field usually requires a numerical approach. Researchers have developed finite element (Shin et al. 2003; Munday et al. 2015) and boundary element (Takahashi and Ghoniem 2008) methods for computing the modulus-mismatch field (which is often referred to as the "image" field). Computation of this field is often expensive and time-consuming. For this reason, many researchers simply neglect the influence of a modulus mismatch in DDD simulations.

Short-range interactions are usually incorporated into the mobility law. Three different approaches have been used. For impenetrable precipitates, dislocation segments are simply pinned with their velocity set to zero at the surface of the precipitates (Mohles 2003; Queyreau et al. 2010). Given the complex nature of precipitate cutting discussed above, penetrable precipitates are more difficult to model. A commonly used approach is to impose a "friction stress," similar to the solute drag-induced friction stress discussed in Sect. 3.2.1, on any dislocation segments that penetrate precipitates (Mohles 2001; Rao et al. 2004; Monnet 2006). The magnitude of the friction stress should depend on the physical processes associated with precipitate cutting and could be selected on the basis of atomic simulations. And finally, Monnet (2015) has recently developed a more rigorous method for incorporating atomistic results into DDD while accounting for thermally activated cutting processes.

### 3.2.4  Grain Boundaries

**Background**

Nearly all DDD studies focus on single crystalline materials, despite the fact that most engineering materials are polycrystalline and grain boundaries are known to influence plastic behavior. The primary reason that grain boundaries are neglected is their complexity; many different types of grain boundaries exist, and they can interact with dislocations in numerous ways. Given that so few researchers have focused their studies on grain boundaries, we will only present a brief sketch of their consideration in the context of DDD.

**Nature of the Physics**

Grain boundaries are usually characterized by the amount and type of misorientation between the two grains forming them; the grains can be twisted with respect to each other, tilted with respect to each other, or a combination of both. Furthermore, depending on the misorientation, there may be a set of *coincidence sites* where the two crystals share a lattice site, or there may be no shared sites at all. In the former case, the grain boundaries are referred to as *singular boundaries* because they have lower energies than more general boundaries without coincidence sites. Singular boundaries are interesting because they can be described in terms of a set of grain boundary dislocations whose Burgers vectors are dictated by the so-called *displacement shift complete* (DSC) lattice of the boundary. Some of these grain boundary dislocations may be mobile within the boundary. When a bulk lattice dislocation encounters a singular grain boundary, it may absorb into the boundary and disassociate into grain boundary dislocations (such that the Burgers vector is conserved). Or the inverse can happen, with grain boundary dislocations transferring into the bulk. In general, it is difficult to quantify the probability for dislocation absorption/emission into/out of a grain boundary, and typically atomistic simulations are necessary to analyze a particular grain boundary of interest (de Koning et al. 2003; Sangid et al. 2011; Spearot and Sangid 2014). In the simplest case, transmission into/across a grain boundary is impossible, which leads to dislocation pileups forming on either side of the boundary. See Cai and Nix (2016) for a more detailed discussion on grain boundaries.

**Incorporation in DDD**

The key rules governing dislocation-grain boundary interactions that must be considered for DDD simulations are as follows: (1) criteria for transmission of a bulk dislocation into a grain boundary, (2) criteria for emission of a grain boundary dislocation into the bulk, and (3) selection of the particular bulk/grain boundary dislocation topology (Burgers vectors, glide planes, etc.) associated with a transmission/emission event. Also, if grain boundary dislocations are glissile, mobility laws must be defined for their motion within the boundary. Strength criteria have been employed by a number of authors to model transmission/emission (Fan et al. 2012; Zhou and LeSar 2012), such that a dislocation may only transmit/emit if

the local shear stress exceeds some minimum. Typically this strength is dictated by the amount of misorientation, and an infinite strength corresponds to an impenetrable grain boundary. Selection of the associated topology is trickier, because the DSC lattice typically has smaller Burgers vectors than the bulk lattice (Cai and Nix 2016), meaning a bulk dislocation has many possible disassociated forms. Criteria have been developed based on energy/dissipation rate considerations similar to those used for junction formation and splintering (see Sects. 2.3 and 3.1.3) (Fan et al. 2012, 2015).

## 4    Examples

### 4.1    Orowan Looping Around Misfitting Platelet Precipitates

Depending upon the heat treatment and alloy composition, aluminum-copper alloys form four different platelet-shaped precipitates. $\theta'$ precipitates are arguably the most important. These precipitates are $Al_2O$ compounds, reside in {001} planes, and are coherent with the FCC Al matrix. However, the lattice mismatch on the faces of the precipitate is different from the mismatch around the rim, leading to a non-dilatational eigenstrain (Biswas et al. 2011). This mismatch makes it difficult for dislocations to transmit into and cut the precipitates. To study Orowan looping around these precipitates, we treat them as impenetrable ellipsoids with two major axes, setting a node's velocity to zero upon contact with the precipitate surface. The lattice-mismatch stress field is pre-computed numerically on a nonuniform grid and then interpolated for Peach-Koehler force calculation (Sills 2016). For simplicity, we neglect the modulus-mismatch field. We compute the shear stress necessary for Orowan looping by starting with a straight edge dislocation and incrementing the applied stress until Orowan looping occurs. An infinite array of parallel, ellipsoidal precipitates with major diameters of 4000 Å and aspect ratios of 30 is considered.

Figure 4a shows the critical dislocation configuration just below the looping stress. The misfit shear stress field is shown as well. Figure 4b shows how the looping stress varies with the spacing between the precipitates. We also show results without a lattice-mismatch field. The lattice-mismatch field is shown not to significantly influence the results, when the dislocation approaches the precipitate array both from the right and from the left.

### 4.2    Junction Splintering in HCP Metals

Discrete dislocation dynamics simulations can add some insights into mechanisms of formation and dissociation of <c + a> dislocations in HCP metals. We consider here the behavior of two reacting <c + a> dislocations. Taking into account the symmetry of the HCP lattice, there are eight distinct types of binary reactions between two <c + a> dislocations (Aubry et al. 2016). Out of those, only four are elastically attractive, and of those four, two reactions lead to the formation of composite dislocations.
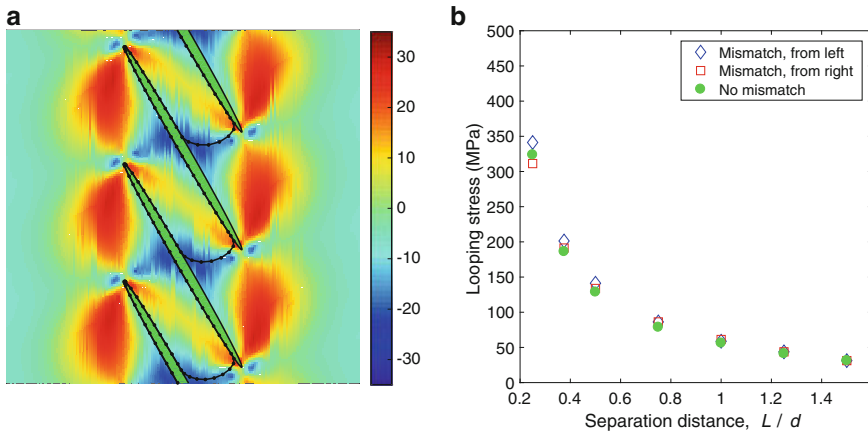
**Fig. 4** Orowan looping results for $\theta'$ precipitates in aluminum-copper. (**a**) Cross section of a precipitate array at the midplane showing the dislocation configuration just below the Orowan looping stress. The color map shows the lattice-mismatch shear stress field in MPa. (**b**) The shear stress necessary for an edge dislocation to Orowan loop as a function of separation distance. $L$ is the separation distance between precipitates and $d$ is their major diameter. The results with and without a lattice-mismatch field are shown. With a mismatch field, we show results when the dislocation approaches from the right and from the left. (Figure adapted from Sills 2016 with permission)



Initial configuration          No dissociation          Dissociation allowed
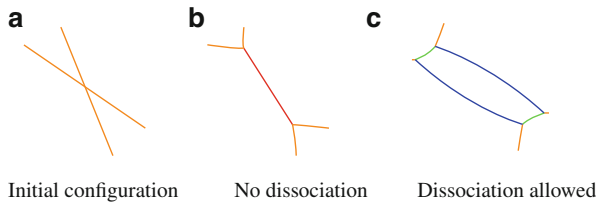
**Fig. 5** DDD simulation of junction formation and splintering between two <c+a>-type dislocations. (**a**) Initial configuration, (**b**) resultant configuration when no composite dislocation dissociation (splintering) occurs, and (**c**) resultant configuration when dissociation of composite dislocations is allowed. Colors denote Burgers vector types: orange = <c+a>, blue = <a>, green = <c>, red = <2a>. (Figure adapted from Aubry et al. 2016 with permission)

Let us analyze a specific example. Consider two parent dislocations with Burgers vectors $\boldsymbol{b}_1 = 1/3[\bar{1}\bar{1}23]$ and $\boldsymbol{b}_2 = 1/3[\bar{1}\bar{1}2\bar{3}]$. The Burgers vector of the intermediate product dislocation formed after their reaction is $\boldsymbol{b}_{ip} = \frac{2}{3}[\bar{1}\bar{1}20]$. Since $\boldsymbol{b}_{ip}$ is twice the Burgers vector of a primary <a> dislocation $\boldsymbol{b}$, the junction should further dissociate into two <a> dislocations. The expected total dislocation reaction can be written as

$$\boldsymbol{b}_1 + \boldsymbol{b}_2 \rightarrow \boldsymbol{b}_{ip} \rightarrow \boldsymbol{b} + \boldsymbol{b}.$$

In Fig. 5a, we consider the collision of two <c+a> dislocations with Burgers vectors $\boldsymbol{b}_1$ set on the pyramidal plane $(01\bar{1}1)$ and $\boldsymbol{b}_2$ set on the prismatic plane $(1\bar{1}00)$. Both parent dislocations form a 40° angle with respect to the junction axis in

their initial orientations. The incidence angle between the intersecting dislocations is such that the dislocation lines are attracted to each other, causing them to react and form a <2a> junction product in the manner shown in Fig. 5b. The resulting <2a> junction dislocation is contained in the same prismatic plane as one of the parents. We then allow the junction to disassociate (splinter) into two <a> primary dislocations. Note that the nodes bounding the junction are then able to split further, leading to a pair of <c> dislocation segments as shown in Fig. 5c. A large-scale study on the effects of junction dissociation on the stress-strain curve was conducted by Aubry et al. (2016).

# 5    Conclusions

In this chapter, we have shown that discrete dislocation dynamics is a diverse, powerful tool for studying dislocation microstructure evolution and the plastic response of materials. However, it is clear that engineering materials are complex and that the inclusion of all relevant physics can be a daunting task. Indeed, many of the DDD algorithms that have been developed are simple and approximate, if only to render them tractable. A DDD formulation with "all of the physics" is simply not practical. As materials modeling continues to mature as a field and our understanding of these physics grows, DDD codes will be expected to incorporate more and more complexity. This will make it even more important that researchers "keep their eyes on the prize" and remember what it is that they are trying to achieve through DDD in the first place.

# References

Agnew S, Horton J, MH Y (2002) Transmission electron microscopy investigation of <c+a> dislocations in mg and $\alpha$-solids solution Mg-Li alloys. Metall Mater Trans A 33A:851

Argon AS (2008) Strengthening mechanisms in crystal plasticity. Oxford University Press, Oxford

Arsenlis A, Cai W, Tang M, Rhee M, Oppelstrup T, Hommes G, Pierce TG, Bulatov VV (2007) Enabling strain hardening simulations with dislocation dynamics. Model Simul Mater Sci Eng 15(6):553–595

Aubry S, Arsenlis A (2013) Use of spherical harmonics for dislocation dynamics in anisotropic elastic media. Model Simul Mater Sci Eng 21(6):065013

Aubry S, Fitzgerald S, Dudarev S, Cai W (2011) Equilibrium shapes of dislocation shear loops in anisotropic -Fe. Model Simul Mater Sci Eng 19:065006

Aubry S, Rhee M, Hommes G, Bulatov V, Arsenlis A (2016) Dislocation dynamics in hexagonal close-packed crystals. J Mech Phys Solids 94:105–126

Bakó B, Clouet E, Dupuy LM, Blétry M (2011) Dislocation dynamics simulations with climb: kinetics of dislocation loop coarsening controlled by bulk diffusion. Philos Mag 91:3173–3191

Barton NR, Bernier JV, Becker R, Arsenlis A, Cavallo R, Marian J, Rhee M, Park HS, Remington BA, Olson RT (2011) A multiscale strength model for extreme loading conditions. J Appl Phys 109(7):073501

Biswas A, Siegel DJ, Wolverton C, Seidman DN (2011) Precipitates in Al-Cu alloys revisited: atom-probe tomographic experiments and first-principles calculations of compositional evolution and interfacial segregation. Acta Mater 59(15):6187–6204

Boyd JD, Nicholson RB (1971) The coarsening behaviour of $\theta''$ and $\theta'$ precipitates in two Al-Cu alloys. Acta Metall 19: 1379–1391

Bulatov VV, Cai W (2006) Computer simulations of dislocations. Oxford University Press, Oxford

Cai W, Bulatov V (2004) Mobility laws in dislocation dynamics simulations. Mat Sci Eng A 387–389:277–281

Cai W, Nix WD (2016) Imperfections in crystalline solids. Cambridge University Press, Cambridge

Cai W, Sills RB, Barnett DM, Nix WD (2014) Modeling a distribution of point defects as misfitting inclusions in stressed solids. J Mech Phys Sol 66:154–171

Chateau JP, Delafosse D, Magnin T (2002) Numerical simulations of hydrogen-dislocation interactions in FCC stainless steels. Acta Mater 50(6):1507–1522

Chen Q, Liu XY, Biner SB (2008) Solute and dislocation junction interactions. Acta Mater 56(13):2937–2947

Cottrell AH, Jaswon MA (1949) Distribution of solute atoms round a slow dislocation. Proc Phys Soc A 62:104–114

Crone CC, Chung PW, Leiter KW, Knap J, Aubry S, Hommes G, Arsenlis A (2014) A multiply parallel implementation of finite element-based discrete dislocation dynamics for arbitrary geometries. Model Simul Mater Sci Eng 22(3):035014

de Koning M, Kurtz RJ, Bulatov VV, Henager CH, Hoagland RG, Cai W, Nomura M (2003) Modeling of dislocation-grain boundary interactions in FCC metals. J Nucl Mater 323:281–289

Delafosse D (2012) Hydrogen effects on the plasticity of face centred cubic (FCC) crystals, chap 9. In: Gangloff RP, Somerday BP (eds) Gaseous hydrogen embrittlement of materials in energy technologies. Mechanisms, modelling and future developments, vol 2. Woodhead Publishing Limited, Cambridge, pp 247–285

Eshelby JD (1961) Elastic inclusions and inhomogeneities, chap 3. In: Sneddon IN, Hill R (eds) Progress in solid mechanics, vol II. North-Holland Publishing Company, Amsterdam, pp 87–140

Fan H, Li Z, Huang M (2012) Toward a further understanding of intermittent plastic responses in the compress single/bicrystalline micropillars. Scr Mater 66:813–816

Fan H, Aubry S, Arsenlis A, El-Awady JA (2015) The role of twinning deformation on the hardening reponse of polycrsytalline magnesium from discrete dislocation dynamics simulations. Acta Mater 92:126–139

Fitzgerald S, Aubry S, Dudarev S, Cai W (2012) Discrete dislocation dynamics simulation of Frank-Read sources in anisotropic $\alpha$-fe. Modell Simul Mater Sci Eng 20:045022

Gao Y, Zhuang Z, Liu ZL, You XC, Zhao XC, Zhang ZH (2011) Investigations of pipe-diffusion-based dislocation climb by discrete dislocation dynamics. Int J Plast 27:1055–1071

Gilbert MR, Queyreau S, Marian J (2011) Stress and temperature dependence of screw dislocation in $\alpha$-Fe by molecular dynamics. Phys Rev B 85:174103

Groh S, Marin EB, Horstemeyer MF, Bammann DJ (2009) Dislocation motion in magnesium: a study by molecular statics and molecular dynamics. Modell Simul Mater Sci Eng 17(7):075009

Gu Y, Xiang Y, Quek SS, Srolovitz DJ (2015) Three-dimensional formulation of dislocation climb. J Mech Phys Sol 83:319–337

Hirth JP, Lothe J (1992) Theory of dislocations. Wiley, New York

Hull D, Bacon DJ (2011) Introduction to dislocations. Butterworth-Heinemann, Oxford

Kirchheim R (2007) Reducing grain boundary, dislocation line and vacancy formation energies by solute segregation II. Experimental evidence and consequences. Acta Mater 55:5139–5148

Kubin LP (2013) Dislocations, mesoscale simulations, and plastic flow. Oxford University Press, Oxford

Kubin LP, Canova G, Condat M, Devincre B, Pontikis V, Bréchet Y (1992) Dislocation microstructures and plastic flow: a 3D simulation. Solid State Phenom 23 & 24:455–472

Lubarda V, Blume J, Needleman A (1993) An analysis of equilibrium dislocation distributions. Acta Metal Mater 41(2):625–642

Marian J, Caro A (2006) Moving dislocations in disordered alloys: connecting continuum and discrete models with atomistic simulations. Phys Rev B 74:024113

Medvedeva NI, Gornostyrev YN, Freeman AJ (2005) Solid solution softening in BCC Mo alloys: effect of transition-metal additions on dislocation structure and mobility. Phys Rev B 72:124107

Mohles V (2001) Orowan process controlled dislocation glide in materials containing incoherent particles. Mat Sci Eng A 309–310:265–269

Mohles V (2003) Superposition of dispersion strengthening and size-mismatch strengthening: computer simulations. Philos Mag Lett 83:9–19

Mohles V (2004) Dislocation dynamics simulations of particle strengthening, chap 17. In: Raabe D, Roters F, Barlat F, Chen LQ (eds) Continuum scale simulation of engineering materials. Wiley-VCH Verlag GmbH & Co., Weinheim, pp 368–388

Mohles V, Fruhstorfer B (2002) Computer simulations of Orowan process controlled dislocation glide in particle arrangements of various randomness. Acta Mater 50:2503–2516

Monnet G (2006) Investigation of precipitation hardening by dislocation dynamics simulations. Philos Mag 86(36):5827–5941

Monnet G (2015) Multiscale modeling of precipitation hardening: application to the FeCr alloys. Acta Mater 95:302–311

Monnet G, Devincre B (2006) Solute friction and forest interaction. Philos Magaz 86(11):1555–1565

Mordehai D, Clouet E, Fivel M, Verdier M (2008) Introducing dislocation climb by bulk diffusion in discrete dislocation dynamics. Philos Mag 88:899–925

Munday LB, Crone JC, Knap J (2015) The role of free surfaces on the formation of prismatic dislocation loops. Scr Mater 103:65–68

Mura T (1987) Micromechanics of defects in solids. Kluwer Academic Publishers, Dordrecht (NL)

Niu X, Luo T, Lu J, Xiang Y (2017) Dislocation climb models from atomistic scheme to dislocation dynamics. J Mech Phys Sol 99:242–258

Nöhring WG, Curtin WA (2017) Dislocation cross-slip in fcc solid solution alloys. Acta Mater 128:135–148

Olmsted DL, Hector LG, Curtin WA, Clifton RJ (2005) Atomistic simulations of dislocation mobility in Al, Ni and Al/Mg alloys. Model Simul Mater Sci Eng 105:371–388

Po G, Cui Y, Rivera D, Cereceda D, Swinburne TD, Marian J, Ghoniem N (2016) A phenomenological dislocation mobility law for BCC metals. Acta Mater 119:123–135

Queyreau S, Monnet G, Devincre B (2010) Orowan strengthening and forest hardening superposition examined by dislocation dynamics simulations. Acta Mater 58(17):5586–5595

Rao SI, Parthasarathy TA, Dimiduk DM, Hazzledine PM (2004) Discrete dislocation simulations of precipitation hardening in superalloys. Philos Mag 84(30):3195–3215

Reed-Hill RE, Abbaschian R (1992) Physical metallurgy principles. PWS-Kent, Boston

Rhee M, Stolken JS, Bulatov VV, Diaz de la Rubia T, Zbib HM, Hirth JP (2001) Dislocation stress fields for dynamic codes using anisotropic elasticity: methodology and analysis. Model Simul Mater Sci Eng A 309–310:288–293

Sangid MD, Ezaz T, Sehitoglu H, Robertson IM (2011) Energy of slip transmission and nucleation at grain boundaries. Acta Mater 59:283–296

Schwarz KW (2003) Local rules for approximating strong dislocation interactions in discrete dislocation dynamics. Model Simul Mater Sci Eng 11:609–625

Shin CS, Fivel MC, Verdier M, Oh KH (2003) Dislocation-impenetrable precipitate interaction: a three-dimensional discrete dislocation dynamics analysis. Philos Mag 83(31–34):3691–3704

Sills RB (2016) Dislocation dynamics of face-centered cubic metals and alloys. PhD thesis, Stanford University

Sills RB, Cai W (2016) Solute drag on perfect and extended dislocations. Philos Mag 96:895–921

Sills RB, Cai W (2018) Free energy change of a dislocation due to a Cottrell atmosphere. Philos Mag 98(16):1491–1510

Sills RB, Kuykendall WP, Aghaei AA, Cai W (2016) Fundamentals of dislocation dynamics simulations, chap 2. In: Weinberger CR, Tucker GJ (eds) Multiscale materials modeling for nanomechanics. Springer, Cham, pp 53–87

Singh CV, Warner DH (2013) An atomistic-based hierarchical multiscale examination of age hardening in an Al-Cu alloy. Metal Mater Trans A: Phys Metal Mater Sci 44:2625–2644

Sofronis P, Birnbaum HK (1994) Mechanics of the hydrogen-dislocation-impurity interactions–I. Increasing shear modulus. J Mech Phys Solids 43(1):49–90

Spearot DE, Sangid MD (2014) Insights on slip transmission at grain boundaries from atomistic simulations. Curr Opin Solid State Mater Sci 18:188–195

Takahashi A, Ghoniem N (2008) A computational method for dislocation-precipitate interaction. J Mech Phys Solids 56(4):1534–1553

Tang M (2005) A lattice based screw-edge dislocation dynamics simulation of body center cubic single crystals, chap 2.22. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht (NL), pp 827–837

Van der Giessen V, Needleman A (1995) Discrete dislocation plasticity: a simple planar model. Model Simul Mater Sci Eng 3(5):689

Van der Giessen E, Needleman A (2005) Discrete dislocation plasticity, chap 3.4. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht (NL), pp 1115–1131

Varvenne C, Leyson GPM, Ghazisaeidi M, Curtin WA (2017) Solute strengthening in random alloys. Acta Mater 124:660–683

Ventelon L, Lüthi B, Clouet E, Proville L, Legrand B, Rodney D, Willaime F (2015) Dislocation core reconstruction induced by cargon segregation in BCC iron. Phys Rev B 91:220102_1–5

Weinberger CR, Cai W (2007) Computing image stress in an elastic cylinder. J Mech Phys Solids 55(10):2027–2054

Weinberger CR, Aubry S, Lee SW, Cai W (2009a) Dislocation dynamics simulations in a cylinder. In: Proceedings of the dislocations 2008 international conference. IOP conference series: materials science and engineering, Dislocations 2008, Hong Kong

Weinberger CR, Aubry S, Lee SW, Nix WD, Cai W (2009b) Modelling dislocations in a free-standing thin film. Model Simul Mater Sci Eng 17:075007–075034

Wen M, Fukuyama S, Yokogawa K (2007) Cross-slip process in FCC nickel with hydrogen in a stacking fault: an atomistic study using the embedded-atom method. Phys Rev B 75:14410_1–4

Weygand D, Friedman LH, Van der Giessen E, Needleman A (2002) Aspects of boundary-value problem solutions with three-dimensional dislocation dynamics. Model Simul Mater Sci Eng 10(4):437–468

Wu Z, Curtin W (2015) The origins of high hardening and low ductility in magnesium. Nature 526:62–67

Yin J, Barnett DM, Cai W (2010) Efficient computation of forces on dislocation segments in anisotropic elasticity. Model Simul Mater Sci Eng 18:045013

Zbib HM, Khraishi TA (2005) Dislocation dynamics, chap 3.3. In: Yip S (ed) Handbook of materials modeling. Springer, Dordrecht/London, pp 1097–1114

Zhou C, LeSar R (2012) Dislocation dynamics simulations of plasticiy in polycrystalline thin films. Int J Plast 30–31:185–201

# Continuum Dislocation Dynamics: Classical Theory and Contemporary Models

# 74

Anter El-Azab and Giacomo Po

## Contents

### Abstract

The continuum theory of dislocation fields is discussed in this chapter with an emphasis on the formulations relevant to infinitesimal deformation of single crystals. Both the classical and contemporary developments are concisely outlined. The classical theory of dislocation fields is introduced first for static and dynamic dislocation configurations, followed by a brief discussion of the shortcomings of

A. El-Azab (✉)
School of Materials Engineering, Purdue University, West Lafayette, IN, USA
e-mail: aelazab@purdue.edu

G. Po
Mechanical and Aerospace Engineering Department, University of California,
Los Angeles, CA, USA
e-mail: gpo@ucla.edu

the classical theory in predicting plasticity of crystals. In this regard, the lack of connection between the evolution of the dislocation field and internal stress state of the crystal is particularly highlighted. The more recent phenomenological and statistically-based formalisms of continuum dislocation dynamics are then introduced. As discussed in the pertinent sections, these formalisms properly connect the evolution of the dislocation fields with the internal stress state in and thus offer frameworks for predicting the plastic behavior of crystals.

# 1    Introduction

The interpretation of plastic deformation of metals in terms of dislocation motion (Orowan 1934; Polanyi 1934; Taylor 1934), together with the subsequent imaging of crystal dislocations by electron microscopy, marked the beginning of the dislocation-based theory of plasticity (Hirth 1985; Hirsch et al. 2006). Since then, the theory has been instrumental in rationalizing many aspects of metal plasticity in terms of the elementary properties of dislocations (Hirth and Lothe 1982; Argon 2008; Kubin 2013). Such a fundamental understanding has permeated through the field of continuum plasticity up to phenomenological theories developed and used at the scale of engineering components (McDowell 2010; Roters et al. 2010).

Despite the fact that most mechanisms of plasticity are known at the single-dislocation level, several features of plastic deformation related to the collective behavior of dislocations have remained hard to understand and predict, such as the phenomena of strain localization, hardening, self-organization, and patterning. Such phenomena were first investigated by simple models such as those developed by Walgraef and Aifantis (1988), before the method of discrete dislocation dynamics simulation was introduced (Lepinoux and Kubin 1987; Ghoniem and Amodeo 1988), eventually leading to elaborate density-based continuum dislocation dynamics approaches; see Sect. 3. The latter theories build upon early attempts to model plasticity through a density-type representation of dislocations, based on the dislocation density tensor introduced by Nye (1953) and Kröner (1981).

The method of discrete dislocation dynamics simulation tracks the trajectory of an ensemble of dislocations evolving under the influence of external stress and their mutual interactions (Kubin 2013; Bulatov and Cai 2006). A typical strain-controlled simulation provides the average dislocation density and stress evolution as a function of imposed strain. The dislocation density corresponding to a given Burgers vector or a slip system can also be determined from the simulation. The method thus provides full diagnostics of dislocation processes during evolution, including cross slip and dislocation annihilation and junction formation. Density-based models of dislocation dynamics, which are termed continuum dislocation dynamics (CDD) in the sequel, have the same goal of tracking the dislocation density evolution and all internal fields. Early relevant formulations used a density tensor to represent dislocations (Mura 1987). However, because this tensor represents a coarse-grained geometric representation of dislocations of all Burgers vectors, it was not possible to close the evolution problem properly and solve it to find the elastic and plastic fields

as a function of the imposed load boundary conditions. More recent theories provide closure approximations which allow the numerical implementation of continuum dislocation-based plasticity (Groma 1997; El-Azab 2000; Acharya 2001; Acharya and Roy 2006; Hochrainer et al. 2009; Xia and El-Azab 2015a). The progress made in the field prospects the possibility that continuum dislocation dynamics can be a viable approach to tackling the problem of self-organization of dislocations and crystal plasticity at the mesoscale (Xia and El-Azab 2015a, b).

This chapter provides a review of CDD approaches to plastic deformation of crystals. Both classical and contemporary models and theories are briefly discussed with a special focus on those falling within the realm of infinitesimal deformation kinematics of crystals. Following this introduction, the classical continuum theory of dislocations is introduced in Sect. 2, and its limitations are discussed. Section 3 introduces the recent attempts at overcoming these limitations.

## 2 Classical Theory of Dislocation Fields

The equations describing crystal deformation, continuum representation of dislocations, and stress fields, along with relevant constitutive relationships, are summarized in this section. In presenting these equations, we adopt the algebra and calculus of Cartesian tensors and denote by boldface symbols that tensors of all orders and adopt the index notation convention for their components (Lubarda 2002). Unless stated otherwise, all field variables are considered to be dependent on the spatial position vector $\mathbf{x} = x_i \mathbf{e}_i$ and time $t$.

In this chapter we use the following notation. The gradient operator $\nabla \equiv \mathbf{e}_i \partial_i(\cdot) \equiv \mathbf{e}_i \partial(\cdot)/\partial x_i$ applied to a scalar function, $f(\mathbf{x})$, a vector field $\mathbf{w}(\mathbf{x})$, and a second-order tensor field $\mathbf{T}(\mathbf{x})$ yields a vector, a second-order tensor, and a third-order tensor, respectively. In this notation $\mathbf{e}_i$, $i = 1, 2, 3$, refer to the base vectors of the Cartesian coordinate system. Throughout this chapter, a left convention is adopted for the gradient operator. We thus have $\nabla f = \partial_i f \mathbf{e}_i = f_{,i} \mathbf{e}_i$, $\nabla \mathbf{w} = \mathbf{e}_i \partial_i(w_j \mathbf{e}_j) = \partial_i w_j \mathbf{e}_i \mathbf{e}_j = w_{j,i} \mathbf{e}_i \mathbf{e}_j$, and $\nabla \mathbf{T} = \mathbf{e}_i \partial_i(T_{jk} \mathbf{e}_j \mathbf{e}_k) = \partial_i T_{jk} \mathbf{e}_i \mathbf{e}_j \mathbf{e}_k = T_{jk,i} \mathbf{e}_i \mathbf{e}_j \mathbf{e}_k$. This notation also defines the comma convention for partial differentiation with respect to spatial coordinates. Finally, the summation over a repeated index is implied. The left convention for the gradient operator implies that the differential change of the function $f$, vector $\mathbf{w}$, and tensor $\mathbf{T}$ over a spatial distance $d\mathbf{x}$ is given by $df = d\mathbf{x} \cdot \nabla f$, $d\mathbf{w} = d\mathbf{x} \cdot \nabla \mathbf{w}$, and $d\mathbf{T} = d\mathbf{x} \cdot \nabla \mathbf{T}$, respectively. Component-wise, these differential quantities are expressed in the forms $df = f_{,i} dx_i$, $dw_j = w_{j,i} dx_i$, and $dT_{jk} = T_{jk,i} dx_i$. Moreover, the curls of a vector field $\mathbf{w}$ and a second-order tensor field $\mathbf{T}$ are defined by $\nabla \times \mathbf{w} = e_{ijk} \partial_j w_k \mathbf{e}_i = e_{ijk} w_{k,j} \mathbf{e}_i$ and $\nabla \times \mathbf{T} = e_{ijk} \partial_j T_{kl} \mathbf{e}_i \mathbf{e}_l = e_{ijk} T_{kl,j} \mathbf{e}_i \mathbf{e}_l$, where $e_{ikl}$ are the components of the antisymmetric permutation tensor. As with the gradient and curl operators, we adopt a left convention of the divergence operator. The divergence of a vector field $\mathbf{w}$ is given by $\nabla \cdot \mathbf{w} = \partial_i w_i = w_{i,i}$ and that of a second-order tensor $\mathbf{T}$ by $\nabla \cdot \mathbf{T} = \partial_i T_{ik} \mathbf{e}_k = T_{ik,i} \mathbf{e}_k$.

## 2.1    Distortion of Dislocated Crystals

As developed in its time-independent form by Kröner (1981) and in its dynamic version by Mura (1963, 1987), Kosevich (1979), and Kossecka and de Wit (1977a, b), the classical theory of dislocation fields aims to determine the elastic state of a crystal containing a prescribed defect distribution. At infinitesimal deformation, the kinematic basis of the theory consists in the additive split of the gradient of the total displacement field $\mathbf{u}^{\mathrm{t}}$ within the crystal, also known as total *distortion* $\boldsymbol{\beta}^{\mathrm{t}}$, which includes the following contributions:

$$\boldsymbol{\beta}^{\mathrm{t}} = \nabla \mathbf{u}^{\mathrm{t}} = \boldsymbol{\beta}^{\circ} + \boldsymbol{\beta}^{\mathrm{e}} + \boldsymbol{\beta}^{\mathrm{p}}. \tag{1}$$

Here $\boldsymbol{\beta}^{\mathrm{p}}$ is the plastic distortion introduced into a crystal by the creation and motion of dislocations, while $\boldsymbol{\beta}^{\mathrm{e}}$ is the elastic distortion generated by them, which is associated with the incompatibility of the plastic distortion, as we shall see in this section (Kröner 1981, 1995, 1996; Kosevich 1962, 1965, 1979; Mura 1963, 1968, 1969; Kossecka 1974, 1975; Kossecka and de Wit 1977a, b). In Eq. (1), we have included the distortion $\boldsymbol{\beta}^{\circ}$ induced by the loading system through mechanical boundary conditions. In component form, Eq. (1) reads $\beta_{ij}^{\mathrm{t}} = \beta_{ij}^{\circ} + \beta_{ij}^{\mathrm{e}} + \beta_{ij}^{\mathrm{p}}$. Both $\boldsymbol{\beta}^{\circ}$ and $\boldsymbol{\beta}^{\mathrm{e}}$ result in dragging the lattice along into an atomically distorted configuration fixed by the plastic distortion, $\boldsymbol{\beta}^{\mathrm{p}}$, and the applied boundary traction (Kröner 1981). Being elastic in nature, they are associated with internal stresses $\boldsymbol{\sigma}^{\circ} = \mathbf{c} : \boldsymbol{\beta}^{\circ}$ and $\boldsymbol{\sigma} = \mathbf{c} : \boldsymbol{\beta}^{\mathrm{e}}$, respectively, so that the total stress within the crystal is

$$\boldsymbol{\sigma}^{\mathrm{t}} = \boldsymbol{\sigma}^{\circ} + \boldsymbol{\sigma} . \tag{2}$$

The distortion $\boldsymbol{\beta}^{\circ}$ is compatible, i.e., it can be expressed as the gradient of a displacement vector $\mathbf{u}^{\circ}$. That is

$$\boldsymbol{\beta}^{\circ} = \nabla \mathbf{u}^{\circ} , \tag{3}$$

which in component form reads $\beta_{ij}^{\circ} = \partial_i u_j^{\circ}$. Equations (1) and (3) imply that the sum $\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathrm{e}} + \boldsymbol{\beta}^{\mathrm{p}}$ is a compatible distortion field derivable from a displacement field, which is denoted by $\mathbf{u}$. The differential form of the compatibility conditions for the distortion fields $\boldsymbol{\beta}^{\circ}$ and $\boldsymbol{\beta}$ can be stated as follows:

$$\nabla \times \boldsymbol{\beta}^{\circ} = \nabla \times \nabla \mathbf{u}^{\circ} = \mathbf{0} \quad \text{and} \quad \nabla \times \boldsymbol{\beta} = \nabla \times \nabla \mathbf{u} = \mathbf{0}, \tag{4}$$

or alternatively

$$\oint_{\mathscr{C}} \mathrm{d}\mathbf{u}^{\circ} = \mathbf{0} \quad \text{and} \quad \oint_{\mathscr{C}} \mathrm{d}\mathbf{u} = \mathbf{0}, \tag{5}$$

where $\nabla \times (\cdot) = \mathbf{e}_i \times \partial_i (\cdot)$ is the curl operator and $\mathscr{C}$ is any closed curve within the (continuum) crystal. The compatibility conditions (4) follow from the orthogonality of the curl and gradient operators, while the form (5) can be reached from (4) by exploiting Stokes theorem. Introduction of this theorem is left for later after the dislocation density tensor and its areal integral are introduced. Selecting the applied traction to be zero implies that the corresponding stress, $\boldsymbol{\sigma}^\circ$, and elastic distortion $\boldsymbol{\beta}^\circ$ are also zero. In such case, the total stress $\boldsymbol{\sigma}^{\mathrm{t}}$ reduces to $\boldsymbol{\sigma}$, the total distortion $\boldsymbol{\beta}^{\mathrm{t}}$ reduces to $\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathrm{e}} + \boldsymbol{\beta}^{\mathrm{p}}$, and the displacement $\mathbf{u}^{\mathrm{t}}$ reduces to $\mathbf{u}$. The condition (4) may then be rewritten in the form

$$\nabla \times \nabla \mathbf{u} = \nabla \times \boldsymbol{\beta} = \nabla \times (\boldsymbol{\beta}^{\mathrm{e}} + \boldsymbol{\beta}^{\mathrm{p}}) = \mathbf{0}. \tag{6}$$

For dynamical problems, the crystal velocity field $\dot{\mathbf{u}}$ is the time derivative of the displacement:

$$\dot{\mathbf{u}} = \partial_t \mathbf{u}, \tag{7}$$

where $\partial_t (\cdot) = \partial (\cdot)/\partial t$. In this case, it can be easily verified that

$$\nabla \dot{\mathbf{u}} = \dot{\boldsymbol{\beta}}, \tag{8}$$

meaning that the rate of the total distortion is given by the gradient of the crystal velocity field.

## 2.2    Tensor Fields of Dislocations

A number of tensor fields connected with dislocations and their motion are now introduced. These include the dislocation density tensor field, the dislocation flux tensor field, and the dislocation velocity and movement tensor fields.

### 2.2.1    The Dislocation Density Tensor Field

Equation (6) can be rearranged in the form

$$\nabla \times \boldsymbol{\beta}^{\mathrm{e}} = -\nabla \times \boldsymbol{\beta}^{\mathrm{p}}. \tag{9}$$

The curl of a distortion field measures its incompatibility. Equation (9) implies that the distortions $\boldsymbol{\beta}^{\mathrm{e}}$ and $\boldsymbol{\beta}^{\mathrm{p}}$ either are simultaneously compatible or have nontrivial curl and are thus simultaneously incompatible. The same equation also implies that the measures of incompatibility of the elastic and plastic distortion are equal and opposite. The dislocation density tensor $\boldsymbol{\alpha}$ is introduced here as that measure. According to Mura (1963, 1968, 1969) and Kossecka and de Wit (1977a, b), $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{\alpha} = \nabla \times \boldsymbol{\beta}^{\mathrm{e}} = -\nabla \times \boldsymbol{\beta}^{\mathrm{p}}. \tag{10}$$

In component form, this reads $\boldsymbol{\alpha}_{ij} = e_{ikl}\partial_k\beta_{lj}^e = e_{ikl}\beta_{lj,k}^e$ or $\boldsymbol{\alpha}_{ij} = -e_{ikl}\partial_k\beta_{lj}^p = -e_{ikl}\beta_{lj,k}^p$. The dislocation density tensor is a field that depends on position $\mathbf{x}$, and its component $\alpha_{ij}$ is defined as the $x_j$ component of the sum of the Burgers vector of all dislocation lines that intersect a unit area normal to the $x_i$ direction, or, alternatively, it is the sum of $x_j$ components of the Burgers vectors times the projected line length in the $x_i$ direction of all dislocations per unit volume. Since the curl of a field is divergence-free, then

$$\nabla \cdot \boldsymbol{\alpha} = \mathbf{0}, \tag{11}$$

where $\nabla\cdot(\cdot) = \mathbf{e}_i\cdot\partial_i(\cdot)$ denotes the divergence operator. In component form, Eq. (11) reads $\partial_i\alpha_{ij} = \alpha_{ij,i} = 0$. This equation is known as the condition of the conservation of the Burgers vector. When this condition is applied to a dislocation line, it simply means that the line cannot end within the crystal.

The interpretation of the dislocation density tensor in terms of crystal dislocations is important in connecting the continuum theory of dislocation fields and the theory for line dislocations. For a family of dislocations of number $\rho^s$ crossing unit area perpendicular to their common line vector $\boldsymbol{\xi}^s$ and having Burgers vector $\mathbf{b}^s$ at a point in the crystal, the dislocation density tensor is given by

$$\boldsymbol{\alpha}^s = \rho^s\boldsymbol{\xi}^s \otimes \mathbf{b}^s, \tag{12}$$

where for two vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \otimes \mathbf{b} = a_ib_j\mathbf{e}_i\mathbf{e}_j$ denotes their tensor product. In component form, the last equation reads $\alpha_{ij}^s = \rho^s\xi_i^s b_j^s$. When multiple families of dislocations are present, the overall dislocation density tensor is the sum of contributions of such families. That is, $\boldsymbol{\alpha} = \sum_s\boldsymbol{\alpha}^s$.

By taking the dot product of equation (10) with an oriented area element $d\mathbf{S} = \mathbf{n}dS = n_idSe_i$, with $\mathbf{n}$ being the unit normal, and integrating that product over an open surface $\mathscr{S}$ bounded by a contour $\mathscr{C}$, it can be shown that

$$\int_{\mathscr{S}} \mathbf{n} \cdot \boldsymbol{\alpha}dS = -\int_{\mathscr{S}} \mathbf{n} \cdot \nabla \times \boldsymbol{\beta}^p dS = -\oint_{\mathscr{C}} d\boldsymbol{\ell} \cdot \boldsymbol{\beta}^p = \mathbf{b}^{\mathscr{C}}. \tag{13}$$

In the above, Stokes theorem has been used to convert the areal integral over $\mathscr{S}$ to a line integral over the bounding contour $\mathscr{C}$. In the above, $\mathbf{n}$, the normal to $\mathscr{S}$, and the directed length element $\boldsymbol{\ell}$ along $\mathscr{C}$ are defined such that if $\mathscr{S}$ shrinks to a small flat element, $\boldsymbol{\ell}$ circles $\mathscr{C}$ around its normal in a right-hand screw sense. In component form, Eq. (13) reads

$$\int_{\mathscr{S}} n_i\alpha_{ij}dS = -\int_{\mathscr{S}} n_ie_{ikl}\beta_{lj,k}^p dS = -\oint_{\mathscr{C}} \beta_{ij}^p d\ell_i = b_j^{\mathscr{C}}. \tag{14}$$

The vector $\mathbf{b}^{\mathscr{C}} = b_j^{\mathscr{C}} \mathbf{e}_j$ is the resultant Burgers vector of all dislocations piercing $\mathscr{S}$ in all directions and contained within $\mathscr{C}$. This vector can be expressed in terms of the elastic distortion by replacing $\nabla \times \boldsymbol{\beta}^{\mathrm{p}}$ with $-\nabla \times \boldsymbol{\beta}^{\mathrm{e}}$ in Eq. (13).

$$\int_{\mathscr{S}} \mathbf{n} \cdot \boldsymbol{\alpha} \mathrm{d}S = \int_{\mathscr{S}} \mathbf{n} \cdot \nabla \times \boldsymbol{\beta}^{\mathrm{e}} \mathrm{d}S = \oint_{\mathscr{C}} \mathrm{d}\boldsymbol{\ell} \cdot \boldsymbol{\beta}^{\mathrm{e}} = \mathbf{b}^{\mathscr{C}}. \qquad (15)$$

It can be shown that the second integral in Eq. (5) is satisfied by subtracting Eq. (13) from Eq. (15) and substituting $\boldsymbol{\beta}^{\mathrm{e}} + \boldsymbol{\beta}^{\mathrm{p}} = \nabla \mathbf{u}$. Equations (10) and (13) are, respectively, known as the differential and integral forms of the incompatibility law. It is obvious that, for a given state of deformation, $\mathbf{b}^{\mathscr{C}}$ vanishes for an arbitrary contour $\mathscr{C}$ if the plastic distortion is curl-free everywhere in the crystal. A uniform plastic distortion satisfies that criterion, in which case it is called a compatible plastic distortion. Such a distortion can be associated with a displacement field satisfying Eq. (5).

### 2.2.2 The Dislocation Velocity Tensor Field

Per Eq. (12), the dislocation density tensor $\boldsymbol{\alpha}$ represents the dislocation state corresponding to the prescribed plastic distortion in the crystal. According to the latter equation, it is possible for the dislocation density tensor to vanish at a point if equal numbers of dislocations of opposite line direction and same Burgers vector pass through that point. From the point of view of plastic deformation, however, all mobile dislocations contribute to the deformation of the crystal regardless of their line direction. As such, to completely characterize moving dislocations and hence the deformation process, different tensors must be introduced that account for the direction of motion of dislocations. One such tensor is the dislocation velocity tensor $\mathbf{V}$, which accounts for the velocity $\mathbf{v}$ of dislocation lines, in addition to their line direction and Burgers vector. Generally, the velocity of a moving dislocation varies from one point to another along its line $\mathscr{L}$. For a family of dislocations of number $\rho^{\mathrm{s}}$ crossing a unit area perpendicular to their common line vector $\boldsymbol{\xi}^{\mathrm{s}}$ and having Burgers vector $\mathbf{b}^{\mathrm{s}}$ at a point in the crystal, Mura (1968, 1969) introduced the dislocation velocity tensor of third rank in the form

$$\mathbf{V}^{\mathrm{s}} = \rho^{\mathrm{s}} \mathbf{v}^{\mathrm{s}} \otimes \boldsymbol{\xi}^{\mathrm{s}} \otimes \mathbf{b}^{\mathrm{s}}. \qquad (16)$$

In component form, the last equation reads $V_{ijk} = \rho^{\mathrm{s}} v_i^{\mathrm{s}} \xi_j^{\mathrm{s}} b_k^{\mathrm{s}}$. When multiple families of dislocations are present, the total dislocation velocity tensor $\mathbf{V}$ is given by the sum of the partial tensors, $\mathbf{V} = \sum_{\mathrm{s}} \mathbf{V}^{\mathrm{s}}$. It is obvious that, when two dislocation families of opposite direction pass at a point, their velocity tensors add together and not cancel each other as do their dislocation density tensors, since the velocities of dislocations of opposite line directions at a point are also opposite. It should be noticed that $V_{mmk} = 0$ since the motion of a dislocation along its line direction is physically insignificant.

### 2.2.3 The Dislocation Movement Tensor Field

The definition of the dislocation movement tensor was given by Eisenberg (1990). Here, this tensor is defined in a slightly different way taking into consideration that it is path dependent and as such can be defined in an incremental sense. The incremental change of the dislocation movement tensor is denoted by d$\mathbf{X}$. For a family of dislocations $\rho^s$ with line tangent $\boldsymbol{\xi}^s$ and Burgers vector $\mathbf{b}^s$, this differential tensor is given by

$$\mathrm{d}\mathbf{X}^s = \mathrm{d}\mathbf{A}^s \otimes \mathbf{b}^s, \tag{17}$$

where d$\mathbf{A}$ is an incremental second-order tensor defined by the outer product of the differential vector distance traversed by dislocations in their glide planes with the oriented dislocation density, $\rho^s \boldsymbol{\xi}^s$. In component form, Eq. (17) reads $\mathrm{d}X^s_{ijk} = \mathrm{d}A^s_{ij} b^s_k$. When multiple families of dislocations are present, $\mathrm{d}\mathbf{X} = \sum_s \mathrm{d}\mathbf{X}^s$. It is clear that

$$\mathbf{V} = \dot{\mathbf{X}} = \partial_t \mathbf{X}, \tag{18}$$

or $\dot{X}_{ijk} = V_{ijk}$. The dislocation movement tensor can thus be obtained from the velocity tensor by a path-dependent integration.

The dislocation velocity and movement tensors carry the overall character of dislocations in terms of their density, line direction, Burgers vector, and velocity vector. Therefore, the velocity tensor can be used to determine the rate of plastic distortion in the crystal containing moving dislocations. Similarly, the incremental change in the dislocation movement tensor can be used to determine the incremental change in the plastic distortion tensor.

### 2.2.4 The Dislocation Flux Tensor Field

A definition of the dislocation flux tensor is given by Kosevich (1979) who wrote Eq. (8) in the form

$$\dot{\boldsymbol{\beta}} = \partial_t \nabla \mathbf{u} = \dot{\boldsymbol{\beta}}^e + \dot{\boldsymbol{\beta}}^p . \tag{19}$$

He then defined the difference $\dot{\boldsymbol{\beta}}^e - \partial_t \nabla \mathbf{u}$ to be the second-order tensor $\mathbf{J}$ and called it the dislocation flux or current tensor. From the last equations, $\mathbf{J} = -\dot{\boldsymbol{\beta}}^p$. In the current presentation, the dislocation flux tensor is defined by

$$\mathbf{J} = \dot{\boldsymbol{\beta}}^p, \tag{20}$$

so that (19) is written as

$$\dot{\boldsymbol{\beta}} = \dot{\boldsymbol{\beta}}^e + \mathbf{J}. \tag{21}$$

By taking the curl of the last equation and using the definition of the dislocation density tensor in (9), one reaches

$$\dot{\boldsymbol{\alpha}} + \nabla \times \mathbf{J} = \mathbf{0}. \tag{22}$$

In component form, Eq. (22) read: $\alpha_{ij} + e_{ikl}\partial_k J_{lj} = 0$. It represents the continuity equation for the dislocation density tensor in a medium undergoing time-dependent plastic distortion. It will be shown later that it can be recast in the form of a transport equation for dislocations. From the definition of the dislocation velocity tensor, it can be verified that the rate of plastic distortion tensor is given by

$$\dot{\beta}_{ij}^{\mathrm{p}} = -e_{imn}V_{mnj}. \tag{23}$$

## 2.3   Dynamics of Dislocated Crystals

The classical development presented so far assumes that the plastic distortion $\boldsymbol{\beta}^{\mathrm{p}}$ and its time rate of change can be introduced arbitrarily, and it is required to determine the displacement, elastic distortion, and internal stress fields. This problem is known as the internal stress or *incompatibility* problem. We continue to present this theory with no regard to the externally imposed stresses as their effect can be simply added to the solution of the incompatibility problem itself.

The elastic distortion $\boldsymbol{\beta}^{\mathrm{e}}$ is related to the internal stress field by Hooke's law for linearly elastic crystals, which is written in the form

$$\boldsymbol{\sigma} = \mathbf{c} : \boldsymbol{\beta}^{\mathrm{e}}, \tag{24}$$

which, in component form, reads $\sigma_{ij} = c_{ijkl}\beta_{kl}^{\mathrm{e}}$, with $\mathbf{c}$ being the fourth-rank elasticity tensor of the crystal. The stress field must satisfy Cauchy's equation of motion which is written as

$$\nabla \cdot \boldsymbol{\sigma} = \rho\ddot{\mathbf{u}}, \tag{25}$$

or $\sigma_{ij,j} = \rho\ddot{u}_i$, where $\rho$ is the mass density of the crystal. For simplicity, we consider the medium to be infinitely extended so that no boundary conditions are included at this stage. Upon using $\boldsymbol{\beta}^{\mathrm{e}} = \boldsymbol{\beta} - \boldsymbol{\beta}^{\mathrm{p}}$ and $\boldsymbol{\beta} = \nabla\mathbf{u}$ into the equation of motion above, the latter can be rearranged in the form

$$\nabla \cdot (\mathbf{c} : \nabla\mathbf{u}) - \rho\ddot{\mathbf{u}} = \nabla \cdot \left(\mathbf{c} : \boldsymbol{\beta}^{\mathrm{p}}\right). \tag{26}$$

or in component form ($c_{ijkl}u_{l,ki} - \rho\ddot{u}_j = c_{ijkl}\beta_{kl,i}^{\mathrm{p}}$). Solution of the equation of motion can be obtained using the Green function technique. Omitting details, the solution for the displacement field is given by Mura (1987),

by

$$u_n(\mathbf{x}, t) = -\int c_{ijkl} G_{jn,i}(\mathbf{R}, T)\beta_{kl}^{\mathrm{p}}(\mathbf{x}', t')\mathrm{d}\mathbf{x}'^3 \mathrm{d}t', \tag{27}$$

and the total distortion tensor by

$$\beta_{mn}(\mathbf{x}, t) = \int c_{ijkl} G_{jn,i}(\mathbf{R}, T)e_{pmk}\alpha_{pl}(\mathbf{x}', t')\mathrm{d}\mathbf{x}'^3 \mathrm{d}t'$$
$$- \int \rho \dot{G}_{ln}(\mathbf{R}, T)J_{ml}(\mathbf{x}', t')\mathrm{d}\mathbf{x}'^3 \mathrm{d}t' + \beta_{mn}^{\mathrm{p}}(\mathbf{x}, t), \tag{28}$$

where $\mathbf{G}(\mathbf{R}, T)$ is the Green function (tensor), $\mathbf{R} = \mathbf{x} - \mathbf{x}'$, $T = t - t'$, and the integrals are carried out over the entire space and time from $-\infty$ to $t$. The elastic distortion tensor is given by $\boldsymbol{\beta}^{\mathrm{e}} = \boldsymbol{\beta} - \boldsymbol{\beta}^{\mathrm{p}}$, and it yields the internal stress via Hooke's law $\sigma_{ij}(\mathbf{x}, t) = c_{ijkl}\beta_{kl}^{\mathrm{e}}(\mathbf{x}, t)$. This completes the solution for the internal stress problem once the dislocation density and flux tensors are prescribed. It can be easily shown that the above results can be specialized to the case of stationary dislocation system.

## 2.4     Summary of the Continuum Dislocation Dynamics Problem

Based on the theory developed in the previous sections, the fields and field equations of continuum dislocation dynamics are summarized as follows:

- *The dislocation field*. The dynamic dislocation field is described by the density tensor $\boldsymbol{\alpha}$ and flux tensor $\mathbf{J}$ or $\boldsymbol{\alpha}$ and the velocity tensor $\mathbf{V}$. The density tensor $\boldsymbol{\alpha}$ and the flux tensor $\mathbf{J}$ are given in terms of the plastic distortion and its time rate of change:

$$\begin{cases} \boldsymbol{\alpha} = -\nabla \times \boldsymbol{\beta}^{\mathrm{p}}, \\ \mathbf{J} = \dot{\boldsymbol{\beta}}^{\mathrm{p}}. \end{cases} \tag{29}$$

These two fields, respectively, satisfy the gauge and continuity conditions

$$\begin{cases} \nabla \cdot \boldsymbol{\alpha} = \mathbf{0}, \\ \dot{\boldsymbol{\alpha}} + \nabla \times \mathbf{J} = \mathbf{0}. \end{cases} \tag{30}$$

- *The distortion fields*. The total distortion $\boldsymbol{\beta}$ is given by the gradient of the displacement field

$$\boldsymbol{\beta} = \nabla \mathbf{u} = \boldsymbol{\beta}^{\mathrm{e}} + \boldsymbol{\beta}^{\mathrm{p}}. \tag{31}$$

The total distortion satisfies the compatibility condition

$$\nabla \times \boldsymbol{\beta} = \mathbf{0}, \tag{32}$$

and the continuity condition

$$\dot{\boldsymbol{\beta}} = \nabla \partial_t \mathbf{u} = \nabla \dot{\mathbf{u}}. \tag{33}$$

• *The displacement and stress fields*. The internal stress field is related to the elastic distortion $\beta$ by Hooke's law

$$\boldsymbol{\sigma} = \mathbf{c} : \boldsymbol{\beta}^{\mathrm{e}}. \tag{34}$$

Therefore the displacement field satisfies Cauchy's equation of motion:

$$\rho \ddot{\mathbf{u}} = \nabla \cdot \boldsymbol{\sigma} = \nabla \cdot \left[ \mathbf{c} : \left( \nabla \mathbf{u} - \boldsymbol{\beta}^{\mathrm{p}} \right) \right] \tag{35}$$

The system of Eqs. (29) through (35) is valid for dynamic problems involving continuously distributed dislocations, and it can be solved provided that the plastic distortion $\boldsymbol{\beta}^{\mathrm{p}}$ and its rate $\dot{\boldsymbol{\beta}}^{\mathrm{p}}$ are prescribed. For a given system of dislocations, the evolution of the dislocation configuration and the plastic distortion can be computed once the dislocation motion is connected with the internal stress field. This closure is missing in the classical theory of dislocation fields. The stumbling block here is the fact that the dislocation density tensor $\boldsymbol{\alpha}$ does not provide sufficient information to determine the rate $\dot{\boldsymbol{\beta}}^{\mathrm{p}}$, which is necessary to evolve the dislocation configuration in time. In fact, the tensor $\boldsymbol{\alpha}$ measures only the so-called *geometrically necessary* dislocations, while $\dot{\boldsymbol{\beta}}^{\mathrm{p}}$ depends in general also on the *statistically stored* dislocation density within a certain material volume. By contrast, in the discrete dislocation dynamics method, all dislocations are fully resolved (in other words they are all geometrically necessary at the length scale of observation), and therefore the evolution of plastic distortion can be determined without complications. The recent continuum dislocation dynamics theories discussed in Sect. 3 achieve this closure of the governing equations and thus qualify as crystal plasticity theories as well.

## 2.5   The Discrete Dislocation Line

The theory presented above can be specialized to discrete dislocation lines. The results can be found in the works by Kosevich (1979), Mura (1987), and Kossecka and de Wit (1977a, b) for the dynamic case and by Kröner (1981) for the static case. Here we summarize the final results for the dynamic case. The term *discrete*

*dislocation line* is used here to represent a single curved dislocation with both ends at infinity, a dislocation line closed on itself, a dislocation network with no free ends in the bulk of the crystal, or a combination of such configurations. The collection of all points on the discrete line will be denoted by $\mathscr{L}(t)$, implying that the dislocation line changes its position and length as a function of time. The first quantity of interest for the discrete dislocation line is the dislocation density tensor field, which is given by

$$\alpha_{pl}(\mathbf{x}, t) = \int_{\mathscr{S}(t)} e_{pmk}\delta_{,m}(\mathbf{R})b_l dS'_k = \oint_{\mathscr{L}(t)} \delta(\mathbf{R})b_l d\ell'_p = \delta_p(\mathscr{L})b_l. \tag{36}$$

where, as previously defined, $\mathbf{R} = \mathbf{x} - \mathbf{x}'$, $\mathscr{S}(t)$ is the defect surface which is bounded by the dislocation line $\mathscr{L}(t)$ and $d\boldsymbol{\ell}'$ is a vector line element at $\mathbf{x}'$. The defect surface $\mathscr{S}(t)$ is the surface of discontinuity of the displacement field. For a glide dislocation loop on a specific crystallographic plane, this surface is enclosed by the loop itself. The function $\delta_p(\mathscr{L}) = \oint_{\mathscr{L}(t)} \delta(\mathbf{R})d\ell'_p$ is a vector Dirac delta function on the line $\mathscr{L}(t)$. Next, the dislocation flux tensor is defined by

$$J_{kl}(\mathbf{x}, t) = \oint_{\mathscr{L}(t)} e_{pmk}\delta(\mathbf{R})b_l v'_m(\mathbf{x}', t)d\ell'_p, \tag{37}$$

where $v_m$ is the $m$th component of the dislocation velocity at $\mathbf{x}'$. The plastic distortion in the case of a discrete dislocation is not a continuous function of space as it was assumed in the case of continuously distributed dislocations. It is the singular part of the total distortion, and it represents a deformation field that is localized on the defect surface $\mathscr{S}(t)$. It is identically zero everywhere else. Its increment or time rate of change, however, is localized on the dislocation line itself as in (37). This plastic distortion field is given by

$$\beta_{kl}^P(\mathbf{x}, t) = -\int_{\mathscr{S}(t)} \delta(\mathbf{R})b_l dS'_k = -\delta_k(\mathscr{S})b_l, \tag{38}$$

where $\delta_k(\mathscr{S})$ is a vector Dirac delta function of the surface $\mathscr{S}$. The elastic distortion for the discrete dislocation line is given by

$$\beta_{mn}(\mathbf{x}, t) = \int dt' \oint_{\mathscr{L}(t')} \left[ e_{pmk}c_{ijkl}G_{jn,i}(\mathbf{R}, T) + \rho\dot{G}_{ln}(\mathbf{R}, T)v'_k(\mathbf{x}') \right] b_l d\ell'_p, \tag{39}$$

and the stress field is given via Hooke's law. The static solution for the incompatibility problem can be obtained by dropping the inertia term in Cauchy's equation of motion and continuing the development in essentially the same way. Such solutions are available in the work of de Wit (1970, 1973).

## 3 Recent Continuum Dislocation Theories

### 3.1 2D Statistical Theory of Dislocations

Groma ([1997]) and Groma et al. ([2003]) pioneered the adoption of the concepts of statistical mechanics to modeling simple, many-dislocation systems, thus developing a first density-based model of dislocation transport in single slip. The dislocation-dislocation correlation was used to account for the impact of dislocation density fluctuations on the internal driving force of dislocations (Zaiser et al. [2001]). Further elaboration of the theory and generalization to multislip case was done by Yefimov et al. ([2004a], [b]) and Yefimov and van der Giessen ([2004], [2005]). Assuming $N$ infinitely long edge dislocations with Burgers vector $b$ distributed in a plane normal to their line direction, the fundamental evolution equation of discrete dislocation dynamics is

$$\mathbf{v}^i = M\mathbf{F}^i_{\mathrm{PK}} = Mb\tau^i_{\mathrm{eff}}\hat{\mathbf{v}}^i\,, \quad \tau^i_{\mathrm{eff}} = \tau^i_{\mathrm{ext}} + \sum_{j\neq i}^N \tau^{ij}(\mathbf{r}^i - \mathbf{r}^j)\,, \quad i, j = 1, \dots, N.$$
(40)

In the above, $\mathbf{v}$ is the dislocation velocity with unit direction $\hat{\mathbf{v}}$, $M$ is the dislocation mobility, $\mathbf{F}_{\mathrm{PK}}$ is the Peach-Koehler force, and $\tau_{\mathrm{eff}}$ is the effective resolved shear stress on the glide plane. The latter stress is split into two parts, the external stress, $\tau_{\mathrm{ext}}$, and the collective stress resulting from interaction with all other dislocations, $\tau^{ij}(\mathbf{r}^i - \mathbf{r}^j)$. A density-based form of the evolution Eq. (40) comes in the form of a pair of transport equations for the density of positively and negatively oriented dislocations, $\rho_+$ and $\rho_-$; see Groma ([1997]) for details. These transport equations have the form

$$\begin{cases} \frac{\partial\rho_+(\mathbf{r},t)}{\partial t} + Mb\frac{\partial}{\partial\mathbf{r}}\int\{\rho_{++}(\mathbf{r},\mathbf{r}',t) - \rho_{+-}(\mathbf{r},\mathbf{r}',t)\}\tau(\mathbf{r} - \mathbf{r}')\mathrm{d}^2\mathbf{r}' \\ \qquad + Mb\frac{\partial}{\partial\mathbf{r}}\rho_+(\mathbf{r},t)\tau_{\mathrm{ext}}(\mathbf{r},t) = 0, \\ \frac{\partial\rho_-(\mathbf{r},t)}{\partial t} + Mb\frac{\partial}{\partial\mathbf{r}}\int\{\rho_{--}(\mathbf{r},\mathbf{r}',t) - \rho_{-+}(\mathbf{r},\mathbf{r}',t)\}\tau(\mathbf{r} - \mathbf{r}')\mathrm{d}^2\mathbf{r}' \\ \qquad + Mb\frac{\partial}{\partial\mathbf{r}}\rho_-(\mathbf{r},t)\tau_{\mathrm{ext}}(\mathbf{r},t) = 0. \end{cases}$$
(41)

In the above, $\rho_{++}(\mathbf{r},\mathbf{r})$, $\rho_{+-}(\mathbf{r},\mathbf{r}')$, $\rho_{--}(\mathbf{r},\mathbf{r}')$, and $\rho_{-+}(\mathbf{r},\mathbf{r}')$ are pair densities, which are cast in terms of the dislocation correlations, $d_{ss'}(\mathbf{r},\mathbf{r}')$, as follows:

$$\rho_{ss'}(\mathbf{r},\mathbf{r}') = \rho_s(\mathbf{r})\rho_{s'}(\mathbf{r}')\left(1 + d_{ss'}(\mathbf{r},\mathbf{r}')\right).$$
(42)

Using these correlations and skipping details, the transport equations can be conveniently rewritten in the form

$$\begin{cases} \frac{\partial\rho(\mathbf{r},t)}{\partial t} + Mb\frac{\partial}{\partial\mathbf{r}}\kappa\{\tau_{\mathrm{int}} + \tau_{\mathrm{ext}} - \tau_{\mathrm{s}}\} = f(\rho, \kappa, \cdots), \\ \frac{\partial\kappa(\mathbf{r},t)}{\partial t} + Mb\frac{\partial}{\partial\mathbf{r}}\rho\{\tau_{\mathrm{int}} + \tau_{\mathrm{ext}} - \tau_{\mathrm{s}}\} = 0, \end{cases}$$
(43)

where $\rho = \rho_+ + \rho_-$ and $\kappa = \rho_+ - \rho_-$ and

$$\tau_{\text{int}}(\mathbf{r}) = \int \kappa(\mathbf{r}')\tau(\mathbf{r}-\mathbf{r}')\text{d}^2\mathbf{r}', \quad \tau_{\text{s}}(\mathbf{r}) = -\int \kappa(\mathbf{r}')d^{\text{t}}(\mathbf{r}-\mathbf{r}')\tau(\mathbf{r}-\mathbf{r}')\text{d}^2\mathbf{r}', \quad (44)$$

with the total correlation, $d^{\text{t}}(\mathbf{r} - \mathbf{r}')$, expressed in the form

$$d^{\text{t}}(\mathbf{r} - \mathbf{r}') = [2d_{++}(\mathbf{r} - \mathbf{r}) + d_{+-}(\mathbf{r} - \mathbf{r}) + d_{+-}(\mathbf{r}' - \mathbf{r})]/4. \quad (45)$$

In the above, $\tau_{\text{int}}$ is the long-range stress, and $\tau_{\text{s}}$ is called the back stress. It is due to the dislocation correlation. Finally, the function $f(\rho, \kappa, \cdots)$ accounts for the nucleation and annihilation of dislocations. Assuming that the correlation is short-ranged and considering an isotropic elastic medium, the back stress can be written as

$$\tau_{\text{s}}(\mathbf{r}) = \frac{\mu b}{2\pi(1 - \nu)\rho(\mathbf{r})} D \frac{\partial \kappa}{\partial \mathbf{r}}, \quad (46)$$

where $\mu$ is the shear modulus, $\nu$ is Poisson's ratio, and $D$ is a dimensionless constant.

The transport equations are solved in conjunction with crystal mechanics equations as shown by Yefimov et al. (2004a, b) and Yefimov and van der Giessen (2004, 2005). In order to do so, the long-range and external stresses are lumped together as the internal stress, which is found by solving the stress boundary value problem:

$$\begin{cases} \nabla \cdot \boldsymbol{\sigma} = \mathbf{0}, & \text{stress equilibrium,} \\ \boldsymbol{\sigma} = \mathbf{c} : \boldsymbol{\varepsilon}^{\text{e}} = \mathbf{c} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\text{p}}), & \text{Hooke's law,} \end{cases} \quad (47)$$

subject to the appropriate traction and/or displacement boundary conditions. In the above, $\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}^{\text{e}}$, and $\boldsymbol{\varepsilon}^{\text{p}}$ are the total, elastic, and plastic strain tensors, respectively. The plastic strain itself is given by

$$\boldsymbol{\varepsilon}^{\text{p}} = \int \dot{\boldsymbol{\varepsilon}}^{\text{p}}(t')dt', \quad \dot{\boldsymbol{\varepsilon}}^{\text{p}} = \dot{\gamma}(\mathbf{S} + \tilde{\mathbf{S}}). \quad (48)$$

where $\mathbf{S} = \mathbf{s} \otimes \mathbf{m}$ is the Schmidt tensor and $\tilde{\mathbf{S}}$ is its transpose, with $\mathbf{s}$ and $\mathbf{m}$ being the unit slip and unit normal vectors, respectively. The plastic shear rate is given in terms of the dislocation transport rate by

$$\dot{\gamma} = \rho b v = \rho b M b(\tau_{\text{RSS}} - \tau_{\text{s}}), \quad \tau_{\text{RSS}} = \boldsymbol{\sigma} : \mathbf{S}. \quad (49)$$

Initial conditions for the dislocation field and the plastic strain are required. The transport equations depend on time explicitly. The mechanics equations depend on time implicitly via $\boldsymbol{\varepsilon}^{\text{p}}(t)$.

Various aspects of this theory have been recently revisited by Valdenaire et al. (2016). Kooiman (2015) on the other hand has developed a more formal coarse-graining procedure of dislocation dynamics of simple, edge-on configurations like the ones treated by Groma (1997) using the General Equation for Non-equilibrium Reversible-Irreversible Coupling (GENERIC) framework (Öttinger 2005).

## 3.2 Field Dislocation Mechanics Theory

Field dislocation mechanics (FDM) is the continuum theory of dislocations developed by Acharya and coworkers (Acharya 2001, 2003, 2004; Roy and Acharya 2005). FDM was formulated with the objective of reconciling conventional plasticity and the elastic theory of continuously distributed dislocations (ECDD) of Willis (1967) and Mura (1987). FDM allows to solve dislocation evolution problems in the fully nonlinear elastoplastic settings. In its early formulation, FDM was intended as a microscopic theory within the finite-deformation kinematics framework, with dislocation density tensors of individual slip systems as primary state variables of the theory. Similar to conventional continuum plasticity, in FDM, the free energy depends only on elastic strains, with no explicit dependence on the dislocation density. Therefore, the theory entails neither higher-order stresses nor a back stress in the expression for average dislocation velocity/plastic strain rate. Subsequent development introduced a new framework named phenomenological mesoscopic field dislocation mechanics (PMFDM) (Acharya and Roy 2006; Roy and Acharya 2006; Acharya 2011), which results from a space-time averaging of the equations of FDM, together with additional phenomenological closure equations. In contrast to FDM, PMFDM is a mesoscale theory, and as such the only microstructural state variable is the average dislocation density tensor field $\boldsymbol{\alpha}$. We shall sketch here a compact summary of PMFDM, limiting our attention to the small deformation case for simplicity.

In broad strokes, the objective of PMFDM simulations is to determine the displacement field $\boldsymbol{u}$ in a certain domain $\Omega$ subject to standard displacement/traction boundary conditions while concurrently evolving the average dislocation density field $\boldsymbol{\alpha}$ subject to dislocation flux boundary conditions. Clearly, the internal stress state is controlled by both the boundary loads and the dislocation density field, and the internal flux of dislocations is driven by the local stress and determines the plastic strain rate.

In PMFDM, a single time step comprises the following three phases:

1. The dislocation density tensor is evolved in time by solving the following IBVP:

$$\begin{cases} \dot{\boldsymbol{\alpha}} = -\text{curl}(\boldsymbol{\alpha} \times \mathbf{v} + \mathbf{L}^{\text{p}}) + \mathbf{s} & \text{in } \Omega \\ \boldsymbol{\alpha}(\mathbf{v} \cdot \mathbf{n}) = \boldsymbol{\phi} & \text{on } \partial_{\text{in}}\Omega \\ \boldsymbol{\alpha}(\mathbf{x}, 0) = \boldsymbol{\alpha}_{\circ}(\mathbf{x}) & \text{in } \Omega, \text{ at } t = 0 \end{cases} \tag{50}$$

Here **v** is the average dislocation velocity, $\mathbf{L}^{\mathrm{p}}$ is the rate of plastic flow carried by the statistically stored dislocations, and **s** is the dislocation nucleation rate. These three quantities are prescribed phenomenologically to provide closure equations to PMFDM. The quantity $\boldsymbol{\phi}$ represents a prescribed flux on the inflow boundary $\partial_{\mathrm{in}}\Omega$, while $\boldsymbol{\alpha}_{\circ}(\boldsymbol{x})$ is the initial condition for the dislocation density. The IBVP (50) is implemented numerically using the so-called Galerkin least-squares (GLS) method. More details on boundary conditions are discussed in Acharya and Roy (2006).

2. Next, two auxiliary fields are determined, namely, the tensor field $\chi$ and the vector field $z$. These two fields are introduced by the *orthogonal decomposition* of the elastic distortion into a compatible part $\nabla(\mathbf{u} - z)$ and an incompatible part $\chi$:

$$\boldsymbol{\beta}^{\mathrm{e}} = \nabla(\mathbf{u} - z) + \chi . \tag{51}$$

The field $\chi$ is determined directly from the dislocation density tensor $\boldsymbol{\alpha}$ by solving the following BVP, say, using least-squares finite element method (LSFEM):

$$\begin{cases} \nabla \times \chi = \boldsymbol{\alpha} & \text{in } \Omega , \\ \nabla \cdot \chi = \mathbf{0} & \text{in } \Omega , \\ \mathbf{n} \cdot \chi = \mathbf{0} & \text{on } \partial\Omega . \end{cases} \tag{52}$$

The vector field $z$ is updated in time using its rate $\dot{z}$, which in turn is found from the solution of the following elliptic BVP:

$$\begin{cases} \nabla \cdot \nabla\dot{z} = \nabla \cdot (\boldsymbol{\alpha} \times \mathbf{v} + \mathbf{L}^{\mathrm{p}}) & \text{in } \Omega \\ \mathbf{n} \cdot \nabla\dot{z} = \mathbf{n} \cdot (\boldsymbol{\alpha} \times \mathbf{v} + \mathbf{L}^{\mathrm{p}}) & \text{on } \partial\Omega \\ \dot{z} = \dot{z}_{\circ} \text{ (arbitrary value)} & \text{at one point in } \Omega. \end{cases} \tag{53}$$

3. Finally, the displacement field is found by a Galerkin FEM implementation of the following BVP:

$$\begin{cases} \nabla \cdot \{ \boldsymbol{C} : [\nabla(\mathbf{u} - z) + \chi] \} = \mathbf{0} & \text{in } \Omega \\ \boldsymbol{\sigma}\mathbf{n} = \mathbf{t} & \text{on } \partial_{\mathbf{t}}\Omega \\ \mathbf{u} = \bar{\mathbf{u}} & \text{on } \partial_{\mathbf{u}}\Omega \end{cases} \tag{54}$$

Recent applications of PMFDM include a study of multicrystalline thin films carried out by Puri et al. (2011). In this study, the mechanical response was
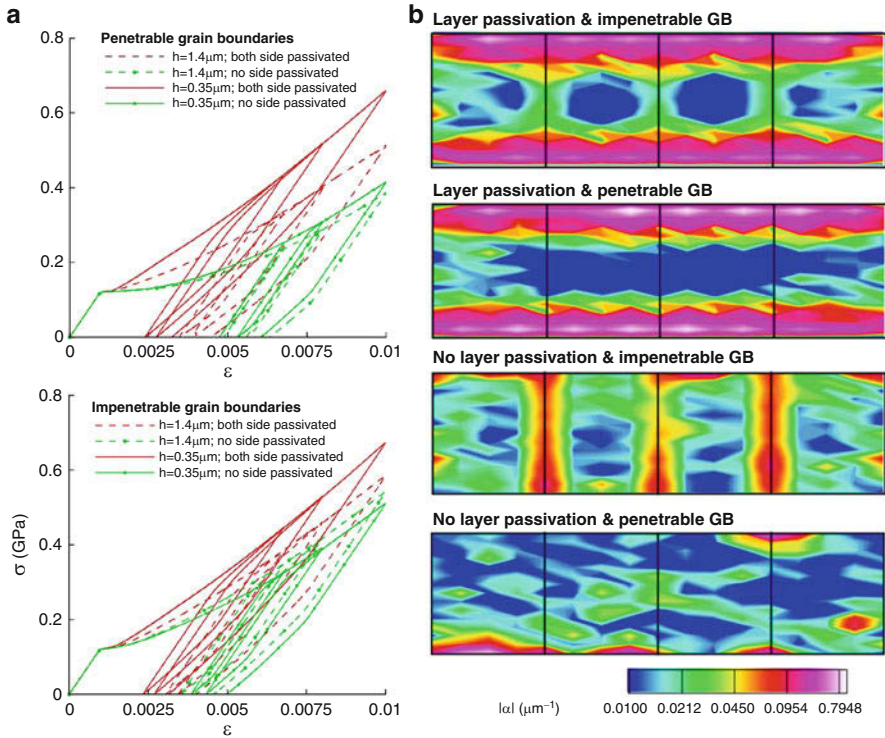
**Fig. 1** Thin film with four grains used in the PMFDM simulations of Puri et al. (2011). (**a**) Stress-strain curves for penetrable and impenetrable grain boundaries. (**b**) Norm of $\boldsymbol{\alpha}$ for different combinations of layer passivation and grain boundary penetrability

analyzed as a function of several parameters including the film thickness, presence of a surface passivation layer, and use of different types of dislocation flux jump conditions to mimic penetrable and impenetrable boundaries. The model predicts that the accumulation of polar dislocations along passivated layers results in a relatively (a) stiffer mechanical response, (b) thickness dependence of stress-strain response, and (c) significant Bauschinger effect in passivated films as compared to unpassivated films, as shown in Fig. 1.

Note that in FDM the dislocation density tensor $\boldsymbol{\alpha}$ ultimately controls its own evolution through the phenomenological closure equation (50). Phenomenological closure equations were also adopted by Sethna and coworkers (Limkumnerd and Sethna 2006; Chen et al. 2010), who introduced a minimal continuum dislocation theory where the evolution of plastic distortion is determined by the traceless portion of the dislocation density as $\dot{\beta}_{ij}^{\mathrm{p}} = v_a \rho_{aij} - 1/3 \delta_{ij} v_a \rho_{akk}$, where $\rho_{ijk} = \beta_{ik,j}^{\mathrm{p}} - \beta_{jk,i}^{\mathrm{p}}$ is simply the Nye tensor expressed as a skew-symmetric tensor of rank three and $v_a$ is a phenomenological dislocation velocity depending on $\rho_{aij}$ itself.

## 3.3 3D Higher-Dimensional Theory

Hochrainer and coworkers (Hochrainer et al. 2007; Sandfeld et al. 2010) developed a continuum theory of dislocations designed specifically to overcome the so-called problem of *kinematic closure*, that is, the problem of defining a set of kinematic variables possessing a sufficient amount of information to determine plastic strain rate and their own kinematic evolution. The theory enjoys kinematic closure in a higher-dimensional configuration space defined by both position and orientation coordinates, hence the name higher-dimensional continuum dislocation dynamics (hdCDD). To illustrate the main ideas, let us consider the two-dimensional case of dislocations with the same Burgers vector $\boldsymbol{b}$ moving on a specific glide plane. Let $\rho^{II}(\boldsymbol{x}, \varphi)$ be the density of dislocations which at the point $\boldsymbol{x} = (x_1, x_2)$ on the plane has a line direction forming an angle $\varphi$ with the Burgers vector. hdCDD defines the generalized line direction as the vector field $\boldsymbol{L}(\boldsymbol{x}, \varphi) = [\cos\varphi, \sin\varphi, k(\boldsymbol{x}, \varphi)]^T$ whose last component is the local line curvature $k(\boldsymbol{x}, \varphi)$. The density $\rho^{II}(\boldsymbol{x}, \varphi)$ and the curvature $k(\boldsymbol{x}, \varphi)$ are the two fundamental fields of hdCDD. Based on these two fields, the dislocation density tensor of second-order $\boldsymbol{\alpha}^{II}$ is defined as

$$\boldsymbol{\alpha}^{II}(\boldsymbol{x}, \varphi) = \rho^{II}(\boldsymbol{x}, \varphi)\boldsymbol{L}(\boldsymbol{x}, \varphi) \otimes \boldsymbol{b}. \tag{55}$$

Because dislocations cannot end within the crystal, this quantity must be divergence-free in the configuration space, which implies the condition

$$\cos\varphi\,\partial_{x_1}\rho^{II} + \sin\varphi\,\partial_{x_2}\rho^{II} + \partial_\varphi(\rho^{II}k) = 0. \tag{56}$$

Next, consider the generalized velocity vector

$$\boldsymbol{V}^{II}(\boldsymbol{x}, \varphi) = [v(\boldsymbol{x}, \varphi)\sin\varphi, -v(\boldsymbol{x}, \varphi)\cos\varphi, \vartheta(\boldsymbol{x}, \varphi)]^T. \tag{57}$$

The first two components of $\boldsymbol{V}^{II}$ represent a velocity field $\boldsymbol{v}(\boldsymbol{x}, \varphi) = v(\boldsymbol{x}, \varphi)[\sin\varphi, -\cos\varphi]$ which is orthogonal to the line direction $[\cos\varphi, \sin\varphi]$ in the glide plane, while the last component is the rotational velocity of the dislocation line. With the understanding that the velocity fields $v(\boldsymbol{x}, \varphi)$ and $\vartheta(\boldsymbol{x}, \varphi)$ are assigned by constitutive laws, it can be shown that the exact time evolution law for $\boldsymbol{\alpha}^{II}$ is

$$\dot{\boldsymbol{\alpha}}^{II}(\boldsymbol{x}, \varphi) = -\nabla \times \left[\boldsymbol{V}^{II}(\boldsymbol{x}, \varphi) \times \boldsymbol{\alpha}^{II}(\boldsymbol{x}, \varphi)\right] \tag{58}$$

Translated in terms of evolution equations for the scalar fields $\rho$ and $k$, Eq. (58) reads:

$$\begin{cases} \dot{\rho}^{II} = -\nabla \cdot (\rho^{II}\boldsymbol{v}) + \partial_\varphi\,(\nabla_{\boldsymbol{L}}v) + \rho^{II}vk \\ \dot{k} = -vk^2 - \nabla_{\boldsymbol{L}}\nabla_{\boldsymbol{L}}v + \nabla_V k \end{cases} \tag{59}$$

In Eq. (59) $\nabla\cdot$ indicates the standard spatial divergence, while $\nabla_X$ indicates the projection of the generalized gradient along direction $X$. Note that the coupling between $\rho$ and $k$ implies the necessity to have curvature as an independent field.

Because all its field equations are defined in the higher-dimensional space $(x, \varphi)$, the numerical implementation of hdCDD has proved to be rather prohibitive. For this reason, subsequent development was aimed at deriving simplified theories based on purely spatial field variables (Hochrainer et al. 2009, 2014; Sandfeld et al. 2011; Hochrainer 2015). The field variables of the simplified theories are the coefficients of the Fourier expansion of the dislocation density and curvature density fields of hdCDD. For example, the zeroth- and first-order coefficients in the expansion of $\rho(x, \varphi)$ are the total scalar dislocation density $\rho_t$ and the geometrically necessary dislocation density vector $\kappa(x)$, respectively:

$$\rho_t(x) = \int_0^{2\pi} \rho(x, \varphi)\,\mathrm{d}\varphi \qquad \kappa(x) = \int_0^{2\pi} \rho(x, \varphi) \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix} \mathrm{d}\varphi. \qquad (60)$$

From $\kappa(x)$, the classical Nye tensor is simply recovered as $\alpha = \kappa(x) \otimes b$. Moreover, under the assumption that scalar dislocation velocity and curvature fields are independent of orientation, averaging of the hdCDD evolution Eq. (59) over $\varphi$ yields the following evolution laws for $\rho_t$ and $\kappa$:

$$\begin{cases} \dot{\rho}_t = -\nabla \cdot (v\kappa^{\perp}) + v\rho_t \overline{k} \\ \dot{\kappa} = -\nabla \times (\rho_t v n) \end{cases} \qquad (61)$$

Here $n$ is the glide plane normal, $\kappa^{\perp} = \kappa \times n$, and $\overline{k}$ is the average curvature. Unfortunately, an evolution law for $\overline{k}$ cannot be expressed in terms of the fields $\rho_t$ and $\kappa$ alone. In fact, the order of truncation of the Fourier expansion defines a hierarchy of theories where the evolution of lower-order Fourier coefficients depends on the values of higher-order ones. Therefore, in contrast to hdCDD, simplified theories lack intrinsic kinematic closure, and phenomenological closure approximations must be introduced at each order. The simplest theory, called $CDD^{(1)}$, is obtained when a phenomenological evolution equation for the average curvature $\overline{k}$ is used together with Eq. (61). A detailed comparison of possible closure equations is discussed by Monavari et al. (2014). This theory was recently applied to study the formation of dislocation patterns by Sandfeld and Zaiser (2015), who found a relationship between the applied stress and the pattern wavelength which is consistent with the principle of similitude, as shown in Fig. 2. Interestingly, in this two-dimensional model, patterns emerge without the cross-slip mechanism.

Despite its importance, it should be remarked that kinematics is only one aspect of the closure problem of CDD discussed in Sect. 2.4. Arguably, the most critical aspect of the closure problem concerns the derivation of the velocity fields in terms of the state variables. Although this process can be cast in a thermodynamically consistent formalism (Acharya 2004; Hochrainer 2016), currently the choice of the thermodynamic potentials remains largely phenomenological.
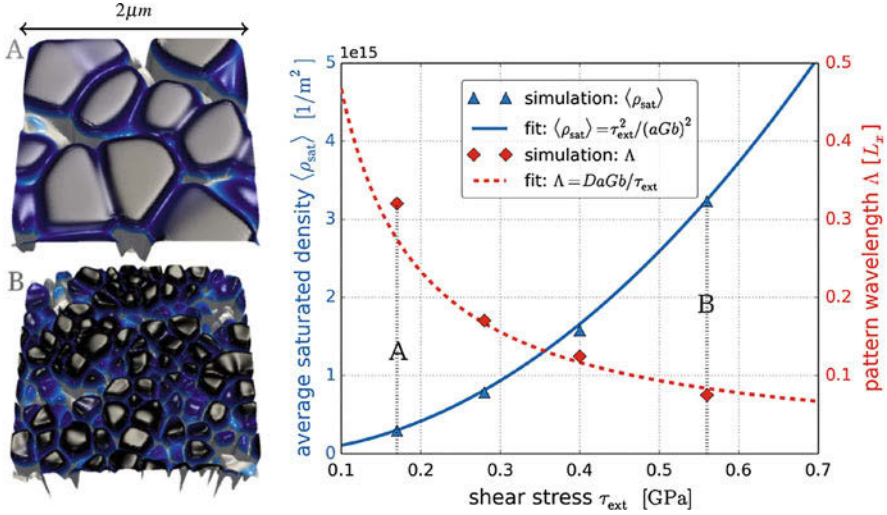
**Fig. 2** Relations between the applied stress $\tau_{ext}$, the patterning length $\Lambda$, and saturated dislocation density $\rho_t$. The evolution of dislocation density patterns in A and B shows a map of the density $\rho_t$ at different levels of $\tau_{ext}$. Plot (**A**) corresponds to a shear stress $\tau_{ext} = 170$ MPa; the plot (**B**) is at $\tau_{ext} = 560$ MPa. (From Sandfeld and Zaiser 2015)

## 3.4    3D Model with a Vector Density Representation

A 3D model of continuum dislocation dynamics was developed by Xia and El-Azab (2015a, b) and Xia et al. (2016). Based on small deformation kinematics, the model considers a vector representation of the dislocation density for the individual slip systems, which are governed by transport-reaction equations derived from the second of Eq. (30) in a few steps: first, that equation is specialized for individual slip systems by adding reaction terms at the level of slip system that cancel one another upon addition and, second, dropping the Burgers vectors from the result. Considering only cross slip as a source terms at slip system level and omitting details, the dislocation transport-reaction equations has the form

$$\dot{\boldsymbol{\rho}}_l = \nabla \times (\mathbf{v}_l \times \boldsymbol{\rho}_l) + \dot{\boldsymbol{\rho}}_{l*\to l}^{cs} - \dot{\boldsymbol{\rho}}_{l\to l*}^{cs} \tag{62}$$

where $\boldsymbol{\rho}_l$ and $\mathbf{v}_l$ are the dislocation density velocity vectors on slip system $l$, respectively, and $\dot{\boldsymbol{\rho}}_{l*\to l}^{cs}$ and $\dot{\boldsymbol{\rho}}_{l\to l*}^{cs}$ are the cross-slip rates from $l*$ to $l$ and from $l$ to $l*$. In the above, $l = 1, \ldots, N$ where $N$ is the total number of slip systems and $l*$ is the cross slip system for system $l$. The above system of equations requires the velocity fields $\mathbf{v}_l$ for all $l$ and the cross-slip rates to be solved. The velocity fields are found in terms of the local Peach-Koehler, $\mathbf{F}_{l,PK}$, via a mobility law

$$\mathbf{v}_l = M\mathbf{F}_{l,PK} = Mb\tau_l\hat{\mathbf{v}}_l \tag{63}$$
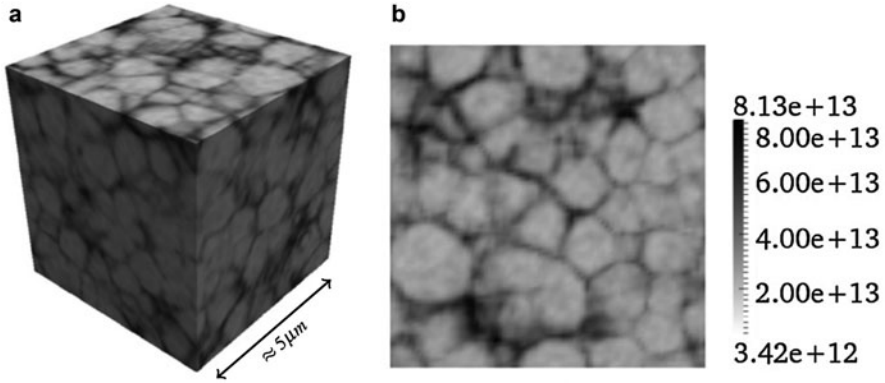
**Fig. 3** (**a**) The 5 μm cube used for the simulation at 0.5% strain, with active cross slip. Colors represent the dislocation density and highlight the 3D cell structure. (**b**) Dislocation structure on a (001) slice. (From Xia et al. 2016)

in which $M$ is the dislocation mobility, $\tau_l$ is the resolved shear stress on slip systems $l$, and $\hat{\mathbf{v}}_l = \mathbf{F}_{l,PK}/|\mathbf{F}_{l,PK}|$ is the unit direction of the velocity. The mobility equation is coupled to the crystal mechanics equations via an eigenstrain problem similar to that summarized in Eq. (47) through (49). Namely, the velocity is found from the stress solution, while the stress boundary value problem is solved by updating the eigenstrain from the dislocation motion via Orowan's law. The cross-slip rate is assumed to have the form

$$\dot{\boldsymbol{\rho}}_{l \to l*}^{cs} = \dot{p}_{l \to l*}^{cs} \boldsymbol{\rho}_{l \| \mathbf{b}}, \tag{64}$$

where $\dot{p}_{l \to l*}^{cs}$ is the cross-slip probability rate and $\boldsymbol{\rho}_{l \| \mathbf{b}}$ is the screw dislocation density. It is important to mention here that the cross-slip rate term is localized in the screw direction only. Xia et al. (2016) has obtained cross-slip rates for an fcc crystal from dislocation dynamics simulations and implemented this rate into the coupled dislocation transport/stress solution outlined above. Figure 3 shows the results of a uniaxial tension simulation in a Cu single crystal. It was found that three-dimensional cell patterns the qualitatively resemble those observed by transmission electron microscopy were obtained when cross slip is activated.

## 4   Summary

The basic concepts of the classical theory of dislocation fields were summarized in this chapter. Of particular interest here is the linking of the dislocation content with the incompatibility of the elastic and plastic distortion fields, which is accomplished by the definition of the dislocation density tensor. Such a tensor was defined on a

purely kinematical basis in the work of Kröner (1981) and de Wit (1970, 1973). Mura (1987) and Kosevich (1979) on the other hand tried to link that tensor to actual crystal dislocations by defining the plastic stain rates in terms of dislocation motion. However, that connection was left incomplete in the classical theory, and as such the theory did not materialize into a framework for predicting plasticity of crystals. Realizing the lack of ability to predict plasticity, and motivated by the method of dislocation dynamics simulation (Kubin 2013; Bulatov and Cai 2006), where the dislocation motion is connected with the internal stress, several authors started to develop predictive continuum dislocation approaches that overcome this shortcoming. The field dislocation mechanics (Acharya 2001, 2003, 2004) is one such approach in which the definition of the dislocation density is linked to the incompatibility equation much as in the classical theory. For other models the need was realized to separate dislocations based on their slip systems and a more detailed description of the transport of dislocations has been used accordingly as in the 2D statistical theory, 3D higher-dimensional theory, and the 3D model with vector representation of the dislocation density. All such approaches are under active development at present time. This development is focused upon three aspects: improving the representation of dislocation mechanisms such as cross slip, junction formation, and annihilation reactions into the transport equations, accounting for long-range stress fluctuations and dislocation-dislocation correlations, and improving the numerical coupling of the transport description of dislocations with crystal mechanics.

# References

Acharya A (2001) A model of crystal plasticity based on the theory of continuously distributed dislocations. J Mech Phys Solids 49(4):761–784

Acharya A (2003) Driving forces and boundary conditions in continuum dislocation mechanics. Proc R Soc Lond A Math Phys Eng Sci 459(2034):1343–1363

Acharya A (2004) Constitutive analysis of finite deformation field dislocation mechanics. J Mech Phys Solids 52(2):301–316

Acharya A (2011) Microcanonical entropy and mesoscale dislocation mechanics and plasticity. J Elast 104(1–2):23–44

Acharya A, Roy A (2006) Size effects and idealized dislocation microstructure at small scales: predictions of a phenomenological model of mesoscopic field dislocation mechanics: part I. J Mech Phys Solids 54(8):1687–1710

Argon A (2008) Strengthening mechanisms in crystal plasticity. Oxford University Press, New York

Bulatov V, Cai W (2006) Computer simulations of dislocation. Oxford University Press, New York

Chen YS, Choi W, Papanikolaou S, Sethna JP (2010) Bending crystals: emergence of fractal dislocation structures. Phys Rev Lett 105(10):105501

de Wit R (1970) Linear theory of static disclinations. In: Simmons J, de Wit R, Bullough R (eds) Fundamental aspects of dislocation theory: National Bureau of Standards special publication 317, vol I. National Bureau of Standards, Washington, pp 651–680

de Wit R (1973) Theory of disclinations: II. Continuous and discrete disclinations in anisotropic elasticity. J Res Natl Bur Stand Phys Chem 77:49–100

Eisenberg M (1990) On viscoplasticity and continuum dislocation theory. In: Weng G, Taya M, Abe H (eds) Micromechanics and inhomogeneity: the Toshio Mura 65th anniversary volume. Springer, New York, pp 115–126

El-Azab A (2000) Statistical mechanics treatment of the evolution of dislocation distributions in single crystals. Phys Rev B 61:11956–11966

Ghoniem NM, Amodeo R (1988) Computer simulation of dislocation pattern formation. Solid State Phenom 3&4:377–388

Groma I (1997) Link between the microscopic and mesoscopic length scale description of the collective behavior of dislocations. Phys Rev B 56:5807–5813

Groma I, Csikor F, Zaiser M (2003) Spatial correlations in higher-order gradient terms in a continuum description of dislocation dynamics. Acta Mater 51:1271–1281

Hirsch P, Cockayne D, Spence J, Whelan M (2006) 50 years of TEM of dislocations: past, present and future. Philos Mag 86:4519–4528

Hirth J (1985) A brief history of dislocation theory. Metall Trans A 16:2085–2090

Hirth J, Lothe J (1982) Theory of dislocations. Wiley, New York

Hochrainer T (2015) Multipole expansion of continuum dislocations dynamics in terms of alignment tensors. Philos Mag 95(12):1321–1367

Hochrainer T (2016) Thermodynamically consistent continuum dislocation dynamics. J Mech Phys Solids 88:12–22

Hochrainer T, Zaiser M, Gumbsch P (2007) A three-dimensional continuum theory of dislocation systems: kinematics and mean-field formulation. Philos Mag 87(8–9): 1261–1282

Hochrainer T, Zaiser M, Gumbsch P (2009) Dislocation transport and line length increase in averaged descriptions of dislocations. AIP Conf Proc 1168(1):1133–1136. https://doi.org/10.1063/1.3241258

Hochrainer T, Sandfeld S, Zaiser M, Gumbsch P (2014) Continuum dislocation dynamics: towards a physical theory of crystal plasticity. J Mech Phys Solids 63:167–178

Kooiman M (2015) Collective dynamics of dislocations. Technical University of Eindhoven, Eindhoven

Kosevich A (1962) The deformation field in an anisotropic elastic medium containing moving dislocations. Sov Phys JETP 15:108–115

Kosevich A (1965) Dynamical theory of dislocations. Sov Phys USPEKHI 7:837–854

Kosevich A (1979) Crystal dislocations and the theory of elasticity. In: Nabarro F (ed) Dislocations in solids, vol 1. North-Holland, Amsterdam, pp 33–141

Kossecka E (1974) Mathematical theory of defects. Part I. Statics. Arch Mech 26:995–1010

Kossecka E (1975) Mathematical theory of defects. Part II. Dynamics. Arch Mech 27:79–92

Kossecka E, de Wit R (1977a) Disclination dynamics. Arch Mech 29:749–767

Kossecka E, de Wit R (1977b) Disclination kinematics. Arch Mech 29:633–651

Kröner E (1981) Continuum theory of defects. In: Balian R, Kléman M, Poirier JP (eds) Continuum theory of defects. North-Holland, Amsterdam, pp 217–315

Kröner E (1995) Dislocation in crystals and in continua: a confrontation. Int J Eng Sci 31: 2127–2135

Kröner E (1996) Dislocation theory as a physical field theory. Meccanica 31:577–587

Kubin L (2013) Dislocations, mesoscale simulations and plastic flow, Oxford Series in Materials Modeling vol. 5. Oxford University Press, Oxford

Lepinoux J, Kubin LP (1987) The dynamic organization of dislocation-structures: a simulation. Scripta Metall Mater 21(6):833–838

Limkumnerd S, Sethna JP (2006) Mesoscale theory of grains and cells: crystal plasticity and coarsening. Phys Rev Lett 96(9):095503

Lubarda V (2002) Elastoplasticity theory. CRC Press LLC, New York

McDowell D (2010) A perspective on trends in multiscale plasticity. Int J Plast 26:1280–1309

Monavari M, Zaiser M, Sandfeld S (2014) Comparison of closure approximations for continuous dislocation dynamics. MRS Online Proceedings Library Archive 1651

Mura T (1963) Continuous distribution of moving dislocations. Philos Mag 8:843–857

Mura T (1968) Continuum theory of dislocations and plasticity. In: Kröner E (ed) Mechanics of generalized continua. Springer, Berlin-Heidelberg, pp 269–278

Mura T (1969) Method of continuously distributed dislocations. In: Mura T (ed) Mathematical theory of dislocations. American Society of Mechanical Engineers, New York, pp 25–48

Mura T (1987) Micromechanics of defects in solids. Matinus-Nijhoff Publishers, Dordrecht

Nye JF (1953) Some geometrical relations in dislocation crystals. Acta Metall 1:153–162

Orowan E (1934) The crystal plasticity. III: about the mechanism of the sliding. Z Physik 89: 634–659

Öttinger HC (2005) Beyond equilibrium thermodynamics. Wiley Interscience Publishers, Hoboken

Polanyi M (1934) Lattice distortion which originates plastic flow. Z Phys 89(9–10):660–662

Puri S, Das A, Acharya A (2011) Mechanical response of multicrystalline thin films in mesoscale field dislocation mechanics. J Mech Phys Solids 59(11):2400–2417

Roters F, Eisenlohr P, Hancherli L, Tjahjanto D, Bieler T, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Int J Plast 58:1152–1211

Roy A, Acharya A (2005) Finite element approximation of field dislocation mechanics. J Mech Phys Solids 53(1):143–170

Roy A, Acharya A (2006) Size effects and idealized dislocation microstructure at small scales: predictions of a phenomenological model of mesoscopic field dislocation mechanics: part II. J Mech Phys Solids 54(8):1711–1743

Sandfeld S, Zaiser M (2015) Pattern formation in a minimal model of continuum dislocation plasticity. Modell Simul Mater Sci Eng 23(6):065005

Sandfeld S, Hochrainer T, Gumbsch P, Zaiser M (2010) Numerical implementation of a 3D continuum theory of dislocation dynamics and application to micro-bending. Philos Mag 90(27–28):3697–3728

Sandfeld S, Hochrainer T, Zaiser M, Gumbsch P (2011) Continuum modeling of dislocation plasticity: theory, numerical implementation, and validation by discrete dislocation simulations. J Mater Res 26(5):623–632

Taylor GI (1934) The mechanism of plastic deformation of crystals. Part I. Theoretical. Proc R Soc Lond Ser A 145(855):362–387

Valdenaire PL, Le Bouar Y, Appolaire B, Finel A (2016) Density-based crystal plasticity: from the discrete to the continuum. Phys Rev B 93:214111

Walgraef D, Aifantis E (1988) Plastic instabilities, dislocation patterns and nonequilibrium phenomena. Res Mech 23:161–195

Willis JR (1967) Second-order effects of dislocations in anisotropic crystals. Int J Eng Sci 5(2):171–190

Xia S, El-Azab A (2015a) A preliminary investigation of dislocation cell structure formation in metals using continuum dislocation dynamics. IOP Conf Ser Mater Sci Eng 89:012053:1–9

Xia S, El-Azab A (2015b) Computational modeling of mesoscale dislocation patterning and plastic deformation of single crystals. Modell Simul Mater Sci Eng 23:055009:1–26

Xia S, Belak J, El-Azab A (2016) The discrete-continuum connection in dislocation dynamics: I. Time coarse graining of cross slip. Modell Simul Mater Sci Eng 24:075007:1–22

Yefimov M, van der Giessen E (2004) Size effect in single crystal thin films: nonlocal crystal plasticity simulations. Eur J Mech A Solids 24:183–193

Yefimov M, van der Giessen E (2005) Multiple slip in a strain-gradient plasticity model motivated by a statistical-mechanics description of dislocations. Int J Solids Struct 42:3375–3394

Yefimov M, Groma I, van der Giessen E (2004a) A comparison of a statistical-mechanics based plasticity model with discrete dislocation plasticity calculations. J Mech Phys Solids 52: 279–300

Yefimov M, van der Giessen E, Groma I (2004b) Bending of a single crystal: discrete dislocations and nonlocal crystal plasticity simulations. Modell Simul Mater Sci Eng 12:1069–1086

Zaiser M, Miguel MC, Groma I (2001) Statistical dynamics of dislocation systems: the influence of dislocation-dislocation correlations. Phys Rev B 64:224102:1–9

# Connecting Lower and Higher Scales in Crystal Plasticity Modeling

# 75

David L. McDowell

## Contents

### Abstract

Metallic materials have a hierarchy of structures ranging in scale from nm to mm. We generalize the notion of crystalline plasticity models to include a range of model constructs that address phenomena associated with evolution of dislocations in crystals across a range of length and timescales. These model constructs range from coarse-grained atomistics, microscopic phase field models, and dislocation field models, to discrete dislocation dynamics, statistical continuum dislocation models, and on up to mesoscale generalized continuum models of gradient, micropolar, or micromorphic type, as well as local continuum crystal plasticity that can be applied over many grains. Key phenomena are introduced and mapped onto the capabilities of various scale-specific model

D. L. McDowell (✉)
Woodruff School of Mechanical Engineering, School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: david.mcdowell@me.gatech.edu

constructs for dislocation plasticity. We discuss concurrent and hierarchical multiscale model transitions in space and time and summarize key challenges in closing.

# 1 Introduction: Material Structure Hierarchy and Associated Models

Metal plasticity is fundamentally associated with processes of nucleation, generation, migration, interaction, trapping, and annihilation of dislocations in crystals and polycrystals (McDowell 2008, 2010). Hierarchy of material structure is prevalent in engineering alloys. Figure 1 depicts five levels of structure hierarchy, ranging from the discrete atomic structure of lattices and interfaces (atomistics), to migration and interaction of dislocation line segments (discrete dislocations), to collective pattern formation of dislocations (substructure), to heterogeneous plastic flow within sets of grains/phases (polycrystal plasticity), and finally up to the scale of engineering applications where underlying structure is "smeared" by considering a macroscopic set of properties or responses. The minimum length scale typical of each of these levels is also shown in Fig. 1 and ranges from interatomic spacing to mean free path
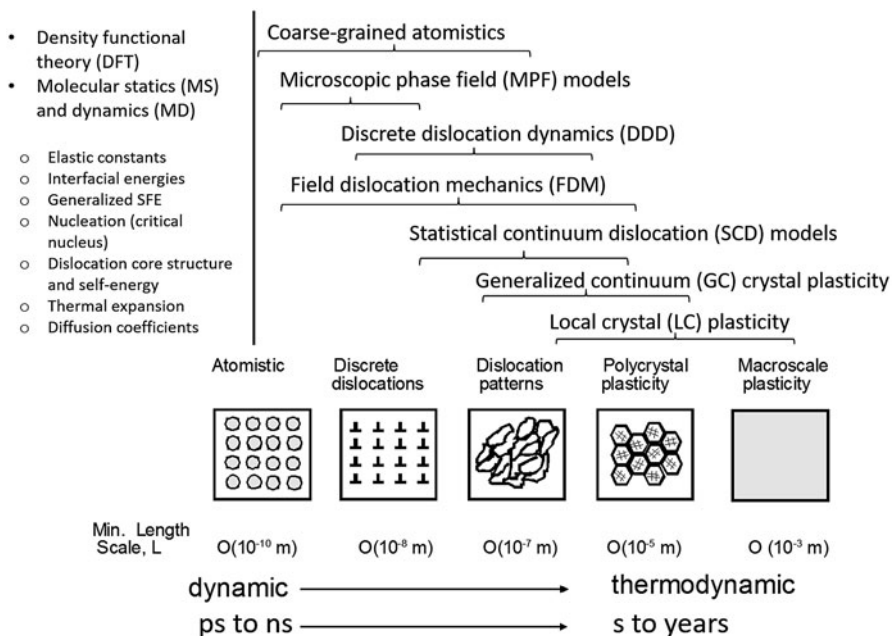


**Fig. 1** Hierarchy of length scales in metal plasticity ranging from atomic (resolution of dislocation cores), to configurations of dislocations, patterning at the mesoscale, and up to multiple grains/phases at the scale of structural applications. Various classes of crystalline plasticity model constructs appear along the top, superimposed on range of associated material structure hierarchy. Certain properties/responses computed using DFT or atomistics shown at far upper left can upscale directly to inform higher length scale models at right

for dislocations, to grain size, and on up to characteristic dimensions of components or structures. Of course, the associated issue of relative timescales of interest for processes at each level of structure hierarchy is implicit in the dynamic to thermodynamic transition indicated from left to right in Fig. 1, ranging from fs to years.

Models of dislocation plasticity are necessary to characterize properties/responses of interest over a wide range of length scales. Goals of this range of model constructs may differ according to the scales addressed. Model constructs addressing phenomena corresponding to scales to the right in Fig. 1 are typically of substantially reduced order and reflect cooperative thermodynamics and kinetics of dislocated crystals. On the other hand, discrete models that apply to scales at the left in Fig. 1 are either fully dynamical or employ some kind of overdamped dynamical scheme, tracking locations of individual defects. Degrees-of-freedom (DOF) necessary to characterize the structure of a fixed volume of material decrease from left to right in Fig. 1. Certain continuum theories are exceptions that bridge to small length scales but potentially retain access to longer timescales relevant to thermally activated dislocation interactions and migration. Examples include microscopic phase field (MPF) models and field dislocation mechanics (FDM) descriptions. FDM descriptions often share common elements with statistical continuum dislocation (SCD) models, namely, incompatibility mechanics based on the net dislocation density tensor, differing mainly in the manner in which distributions of dislocations are tracked. Only atomistic and coarse-grained atomistic approaches commonly address dislocations as naturally emergent line defects in the lattice. With the exception of atomistics (MS or MD), coarse-grained atomistics, or the application of MPF or FDM models with near atomic resolution, all other models introduce primitive representations of dislocations.

## 2 Hierarchical and Concurrent Multiscale Modeling of Dislocations

*For the present purposes, we define multiscale modeling as the practice of building and exercising models for a set of physical processes that operate over a range of length and/or time scales, regardless of the form of linkage between models expressed at various levels of fidelity or resolution.* We may distinguish between coarse-graining and model reduction approaches to multiscale/multiresolution modeling. Here we adopt the most common definition in the discrete particle/molecular modeling communities: coarse-graining reduces the number of model degrees of freedom via simplifications that retain the underlying nature of the original description but average over space and/or time. For example, a model might replace a large number of atoms in a molecular system using a simplified representation while retaining information related to bonding. On the other hand, model reduction (sometimes referred to as model order reduction) seeks to reduce model complexity/fidelity by reducing the model's associated state space dimension or degrees of freedom, offering an approximation to the original model. The key difference is that model form often changes in model reduction, whereas model

form is essentially invariant in coarse-graining. Suffice it to say that most models presented in Fig. 1 involve model reduction rather than coarse-graining; this greatly complicates pursuit of concurrent multiscale modeling approaches, discussed later in this section.

Fully concurrent models seek to simultaneously pursue solutions of models framed at different spatial resolution and corresponding model fidelity (Ghosh et al. 2001, 2007; Kouznetsova et al. 2004). The simulations carried out at different scales are coupled and interact in two-way fashion. They are by nature temporally synchronized, in that the time/strain step is limited by the most demanding resolution of any domain, typically that of highest fidelity and/or resolution. Concurrent models have at least two further classifications: (i) domain decomposition and (ii) multiresolution and/or multifidelity models exercised within or over the same spatial domain. In domain decomposition, a coarse-grained or reduced order model representation is employed over part of a spatial domain, with an abutting high fidelity representation over the remainder. For example, we may require high resolution and fidelity near an interface or a crack tip, with low fidelity and/or resolution elsewhere. For purposes of modeling dislocations, we will consider that fully concurrent models require exchange/transport of dislocations between the two domains; weaker forms of concurrency might consider dislocation migration in only one domain.

Hierarchical multiscale models are of one-way character, typically informing parameters of models at the next successive length scale of modeling within the hierarchy shown in Fig. 1. As shown to far left in Fig. 1, there are some exceptions. Certain properties computed using DFT or atomistics can upscale hierarchically to inform models at various length scales, even up the to the grain scale. Examples include elastic lattice constants, interface energy, generalized stacking fault energy (SFE), and effects of the dislocation core structure and constriction on stress-state dependent yielding.

## 3    Challenges of Dislocation Field Problems: Length and Timescale Transitions

Dislocation field problems in mechanics are particularly demanding for multiscale models, owing to the complexity of the physics (e.g., dislocation core effects, long range dislocation interactions, and short range shielding interactions). Progress has been slow and elusive on concurrent models in this regard, and no satisfactory concurrent multiscale models exist that bridge length scales from dislocation cores (order of nm) to dislocation substructures (hundreds of nm) to typical laboratory specimens ($\mu$m to mm).

There are also substantial challenges in timescaling. Short range lattice relaxation occurs on the order of tens of ps in a MD simulation. Approaches have been introduced to improve time scaling of both direct type (Kim et al. 2014; Binder et al. 2015) or statistical coarse-graining of short range interactions in cross slip of dislocations (cf. Deng and El-Azab 2010). At low applied stress levels, thermally activated bypass of larger obstacles or a field of obstacles with high

activation volume (much larger than scales of atomistic unit processes) requires coordinated thermal fluctuations having low probability of successful attempts. Activation entropy associated with dislocation nucleation and bypass phenomena has been largely neglected. Some works (Nguyen et al. 2011; Ryu et al. 2011; Saroukhani et al. 2016) have considered the limitations of harmonic transition state theory (HTST), since it does not represent anharmonic entropic effects at finite temperature, including temperature-dependence of the elastic shear modulus (thermal softening), thermal expansion, and other surface energies (e.g., stacking fault energy) that relate to dislocation nucleation and bypass.

## 4   Status of Models at Various Levels of Hierarchy

### 4.1   Concurrent Multiscale Modeling Methods from Atomistic to Continuum (QC, CADD, and CAC)

Large scale direct MD simulations of dislocations are highly intensive (e.g., 500 million atom simulations of a slab of Cu by Buehler et al. 2005) and are limited to short times and high strain rates. Historically, concurrent schemes for bridging atomistic and continuum methods in problems involving dislocation activity have employed spatial domain decomposition; fully resolved atomistic models (far left in Fig. 1) are desired in a subdomain near a crack tip or a grain boundary interface, for example, and coarse-grained or reduced order descriptions are used elsewhere. A number of these methods have been compared by Tadmor and Miller (2011).

We focus here on methods that admit moving dislocations in both fully resolved and coarse-grained or continuum regions, as well as two-way exchange of dislocations between the two regions. Clearly, the value of methods that bridge from atomistic to continuum descriptions is their strong predictive character (limited by the uncertainty in the interatomic potential and scale linking algorithms) compared to mesoscale strategies. One of the most widely cited atomistic-continuum multiscaling approaches is the Quasicontinuum (QC) method (Tadmor et al. 1996a, b; Shenoy et al. 1998, 1999; Miller et al. 1998a, b; Knap and Ortiz 2001; Tadmor and Miller 2011). QC accomplishes atomistic coarse-graining by linking lattice site positions to a reduced set of representative atoms, introducing summation rules that allow reduction of cost of ensemble sampling for force or energy, and schemes for adaptation that facilitate fully resolved atomistic simulation in regions of high local deformation (e.g., interfaces, crack tips, dislocations) but coarse-grain away from these evolving regions to minimize the degrees of freedom. In its earliest and perhaps most commonly used form, termed "local" QC, the coarse-grained continuum region makes use of a Cauchy-Born assumption of an affine deformation within each region of interpolation between representative atoms that define elements. Accordingly, it offers a precise description of the underlying interatomic hyperelastic potential even in continuum regions, subject to coarse-graining error for inhomogeneous deformation. QC requires either a priori assignment of fully atomistic and coarse-grained regions for dislocation migration

or adaptation/reassignment of associated atoms and nodes to reach a fully atomistic description near dislocations (Shenoy et al. 1999) or interfaces.

Most applications in modeling dislocations have employed quasi-static QC methods based on energy minimization. A fully nonlocal or cluster QC version was developed (Knap and Ortiz 2001) to minimize ghost forces in coarse-graining, with extensions to energy-based rather than force-based cluster sampling schemes (Eidel and Stukowski 2009. Kwon et al. (2009) have introduced higher order adaptive QC schemes to improve description of large scale inhomogeneous deformation as might occur in coarse-grained regions away from dislocations. QC has been applied to quantify lattice dislocation interactions with grain boundaries (Shimokawa et al. 2007; Yu and Wang 2012, 2014). Recent advances have sought to increase efficiency while handling arbitrary fields of dislocations (Amelang et al. 2013, 2015), as well as development of a fully nonlocal, energy-based QC method by Amelang et al. (2015). This approach has been extended recently to consider automatically adaptive modeling of fully nonlocal QC to model dislocations moving through a lattice (Tembhekar et al. 2017). A meshless QC method based on local maximum-entropy interpolation has been developed (Kochmann and Venturini 2014) that improves efficiency of model adaptation near regions of fully atomistic resolution to capture dislocations.

The Coupled Atomistic Discrete Dislocation (CADD) method introduced originally by Shilkrot et al. (2002a, b, 2004) with subsequent advances and applications (Shiari et al. 2005; Dewald and Curtin 2007a, b, 2011) employs domain decomposition with a continuum discrete dislocation representation away from the grain boundary, admitting two-way exchange of dislocations between this remote region and a fully atomistically resolved domain near the grain boundary. This kind of formulation is quite useful in studies of crack tip and interface plasticity that extends the fields to long range. It is considered fully concurrent in that dislocations can move from the atomistic to continuum regions and vice versa. With a focus on capturing the interaction stress fields of dislocations at long range, Dewald and Curtin (2007a, b) employed CADD to analyze dislocation pile-up interactions with grain boundaries.

Unlike QC, CADD uses a different elastic energy function in the remote regions – these regions are subjected to reduced order continuum models rather than atomistic coarse-graining. Accordingly, in addition to model reduction error, interpretation of temperature differs in the two regions and there is impedance to wave propagation at the atomistic-continuum interface. Impedance effects can be minimized by employing damping heuristics (Qu et al. 2005) but are inevitable in the absence of statistical coarse-graining. Initial implementations of CADD have focused on 2D problems by virtue of complexities in addressing dislocation transfer across interfaces. Progress has been achieved on passing general 3D mixed character dislocations between atomistic and continuum regions (Cho et al. 2015); Pavia and Curtin (2015) have advanced parallel algorithms for 3D CADD that enhance scalability at high accuracy for appropriate problems with long range fields, including Langevin graded stadium damping to eliminate spurious wave reflections and interface forces.

The concurrent atomistic-continuum (CAC) method was developed as an outgrowth of atomistic field theory (Chen 2009), with a local density function used to connect this continuum theory to atomistics via multiscale balance equations. The CAC method is a 3D integral finite element approach to solve the governing field equations for coarse-grained atomistics that admits description of dislocation nucleation and migration with or without adaptive coarse-graining (Xiong et al. 2011, 2012a, b, 2015). Dislocations can be modeled throughout the entire domain, whether at full atomistic resolution or coarse-grained. This sets it apart from methods that require full atomistic resolution at the dislocation core. While QC typically seeks the most accurate and efficient solution to dislocation plasticity via adaptive remeshing of the domain near dislocations to full atomistic resolution, CAC can resolve full atomistics if necessary near interfaces or crack tips, but allows dislocations to nucleate, multiply, migrate, and interact even in the coarse-grained regions along interfaces between elements, introducing the option to coarse-grain dislocation fields over larger scales. Like QC, CAC employs the same interatomic potential even in coarse-grained regions where dislocations evolve and interact. Such coarse-graining in CAC introduces systematic error, for example, in representation of dislocation core structures and short range interactions, which can be quantified and balanced with the high computational demands of remeshing, according to the purposes of the mesoscale modeling.

CAC employs either a quasi-static implementation enforcing quenched dynamics at each time step with periodic energy minimization applied every 50–100 time steps (Xu et al. 2015, 2016), or a true dynamic (full inertial effects that can address phonon interactions with wavelengths above the element size, cf. Xiong et al. 2014a, b, 2015, Pluchino et al. 2016) implementation, based on an extension of the lattice statistical mechanics approaches of Irving and Kirkwood (Kirkwood 1946; Irving and Kirkwood 1950). The quasi-static implementation is considered useful for modeling reaction pathways for thermally activated dislocation processes in a manner that is not overdriven, as typical of MD simulations. Finite temperature applications to date have assumed isothermal conditions or dynamic behavior without a thermostat. Finite temperature formulations for quasi-static (near equilibrium) and nonequilibrium dynamic conditions or heat transport via modes other than phonons remain to be fully developed (cf. QC advances by Marian et al. 2010).

Similar to CADD and static QC, the quasi-static version of CAC is useful for modeling phenomena such as detailed reactions of successive pile-up dislocations with interfaces, with the additional advantage of accessing large volumes that approach scales of TEM thin foils and nonperiodic boundary conditions (e.g., traction free surfaces), cf. Xu et al. 2017.

## 4.2  Microscopic Phase Field (MPF) Models

Continuum microscopic phase field (MPF) models (Wang et al. 2001; Shen and Wang 2003; Wang and Li 2010; Shen et al. 2014) hold significant promise to bridge from first principles and atomistic modeling to describe partial dislocations and deal

with time dependent deformation phenomena in continuum theory in a way that is not limited to the time step size of atomistic simulations. MPF models can be regarded as an application of phase field theory at or near atomic scale of refinement, including diffuse interface theory and gradient terms specific to that scale. MPF has spatial resolution at the core scale, similar to the Peierls model (Peierls 1940), and admits conveyance of information regarding core size, energy and interactions (Chen et al. 2014). Therefore, MPF is framed at a length scale that can resolve individual dislocation cores (i.e., dissociated partial dislocations) via the generalized stacking fault energy function, representing dislocations using eigenstrain methods. It can address limited numbers of dislocations with near atomic scale fidelity. It can directly employ inputs from ab initio and atomistic modeling regarding physical properties, surface energies, and elastic constants, and has the ability to predict fundamental properties of individual defects or small sets of defects, including formation energy, size, structure, saddle point configuration of the reaction pathway, and activation energy of defect nuclei, and the micromechanisms involved (Wang and Li 2010).

When used in conjunction with transition state theory or other appropriate long time kinetics, including diffusive phenomena, continuum phase field theory offers the potential to bridge to long time scales. This is quite useful in predictive computational materials science to support design of interfaces, microstructures, etc. MPF serves to bridge from atomistic modeling of generalized stacking fault energy to describe partial dislocations in a higher scale continuum theory (Wang and Li 2010).

## 4.3 Field Dislocation Mechanics (FDM)

Field dislocation mechanics applies to a theoretical continuum framework that incorporates incompatibility of elastic and plastic deformation as an explicit part of the field equations. Focus is placed on the incompatible part of the plastic distortion that establishes the excess dislocation density tensors. Accordingly, it offers a continuum coarse-graining methodology for incompatibility mechanics by changing the characteristic size of the Burgers circuit (e.g., integration point area through which dislocations cut). The theoretical construct of FDM was formalized by Acharya and co-workers over a decade ago (Acharya 2001, 2003; Acharya et al. 2006, 2008; Roy and Acharya 2006; Roy et al. 2007), providing common ground for treating incompatibilities with higher scale theories such as SCD and GC. Hence, FDM serves as a theoretical construct to consider the effect of size of Burgers circuit on response. Typically based on small strain/distortion assumptions to facilitate solution strategies, this class of theory explicitly incorporates lattice closure relations on incompatibility expressed either in terms of a single dislocation (e.g., approaching atomic scale, cf. Zhang 2015; Hartley and Mishin 2005; Mendis et al. 2006), or in terms of the Nye tensor (Nye 1953) evaluated over a finite Burgers circuit for a large number of dislocations (analogous to second gradient plasticity).

## 4.4    Discrete Dislocation Dynamics (DDD) Models

Discrete dislocation dynamics (DDD) implementations employ continuum linear elasticity solutions for dislocations to compute long range, nonlocal elastic interactions, representing the positions and velocities of all dislocations by discretizing the dislocation lines into a piecewise series of curves or segments (Lepinoux and Kubin 1987; Gulluoglu et al. 1989; Amodeo and Ghoniem 1988, 1990; Ghoniem 2005). DDD methods typically assume a simple form of dislocation force-velocity relationship (mobility) for screw and edge components of dislocations and do not explicitly distinguish partial dislocations. The influence of nonlinear dislocation core effects (which are outside the singular theory of linear elasticity) on short range dislocation interactions has been effectively addressed by non-singular formulations of DDD (Cai et al. 2006) that can replicate the influence of core spreading and attenuation of interaction energies. These approaches can be hierarchically informed using atomistic simulations (Martınez et al. 2008; Groh et al. 2009; Rudd et al. 2014).

The DDD method has received significant investment and is maturing as a tool for discrete dislocation simulations (e.g., ParaDiS code, Arsenlis et al. 2007, Tang et al. 2011; and the microMegas code, Devincre et al. 2011), including massive parallelization (Wang et al. 2006), fast multipole methods to address long range stress field interactions of segments, along with mesh adaptivity to segment the evolving dislocation lines, and employment of maximum power dissipation to assess dislocation junction node dissociation. Sills et al. (2016) have introduced advanced time integration algorithms for DDD simulations. More efficient FFT formulations of DDD have recently been advanced that offer improved efficiency and scalabilty (cf. Bertin et al. 2015). While DDD has been extensively applied to the problem of dislocation pattern formation and workhardening in single crystals under periodic boundary conditions, recent applications have considered the complexity of cyclic loading (Hussein and El-Awady 2016) for up to 80 cycles, including the role of dislocation cell walls.

DDD modeling approaches for finite size specimens have been advanced to consider size effects, influence of free surfaces and image forces, intermittent/jerky flow behavior, and source limitation/starvation phenomena at submicron to micron scale specimens (Van der Giessen and Needleman 1995; Guruprasad and Benzerga 2008; Akarapu et al. 2010). Recent advances have considered highly parallel finite element or boundary element method implementations for arbitrary finite geometries (Crone et al. 2014) and for considering elastically anisotropic crystals (Liu et al. 2016), including associated fast multipole expansions for long range interactions (Yin et al. 2012). DDD has proven extremely useful as a means to model dislocation plasticity in confined geometries such as constrained thin films, multilayers, micropillars, and nanoindentation to provide key physical insights into scale dependent dislocation phenomena and interactions with surfaces and interfaces (cf. Zbib et al. 2011) including cases where kinetics dominates elastic interactions (Ryu et al. 2013).

As a reduced order model, DDD shares certain challenges with other continuum formulations, including construction of high angle grain boundaries and treatment of dislocation slip transfer reactions with interfaces. Representation of solute distribution, associated dislocation core interactions, and drag effects, as in dynamic strain aging, is challenging with DDD and may require two-way multiscale concurrency in cases where timescales of atomic level processes cannot be separated from scales of dislocation migration. Dislocation interactions with point defects (Cai et al. 2014) or interactions with diffusive vacancies at large strain or under cyclic loading at finite temperature require extension of the framework that may vary according to modeler preference and are in early stages of development, whether involving DDD or atomistic methods that separate timescales to inform DDD (e.g., Keralavarma and Benzerga 2015; Baker and Curtin 2016; Matveev et al. 2016).

Certain thermally activated unit processes such as dislocation cross slip have been incorporated heuristically in DDD (cf. Groh and Zbib 2009; Groh et al. 2009; Deng and El-Azab 2010; Kang et al. 2014; Xia and El-Azab 2015; Xia et al. 2016; Hussein and El-Awady 2016). However, just as in coarse-grained atomistics, the ability to model coordinated thermally activated processes involving substantial activation volume (e.g., >1000 b$^3$) is limited.

In terms of two-scale transitions to the continuum crystal plasticity level, the kinematics of DDD are logically addressed by concurrent schemes advanced by Zbib and coworkers (Zbib and Diaz de la Rubia 2002; Zbib et al. 2002; 2011), including coupling with FE methods for addressing finite geometries and consideration of free surfaces (Khraishi and Zbib 2002). DDD has been used to hierarchically inform GC or LC crystal plasticity formulations (cf. Zbib et al. 1998, 2002; Zbib and Diaz de la Rubia 2002; Wallin et al. 2008, Groh et al. 2009). Accounting for continuum lattice rotation in DDD is nontrivial and of multiscale character, requiring an assignment of a volume element for this purpose in multiscale strategies (Zbib and Diaz de la Rubia 2002; Zbib et al. 2002. DDD has also been applied to initial-boundary value problems on finite domains using implementations that couple with the finite element method to model responses of structures and devices (Yasin et al. 2001; Zbib et al. 2011). Groh and Zbib (2009) provide an overview of hierarchical versus concurrent coupling methods between DDD and continuum crystal plasticity.

## 4.5    Statistical Continuum Dislocation (SCD) Models

Statistical continuum dislocation models are a class of reduced order continuum descriptions in the spirit of DDD that formulate the governing equations in a manner that pertains to a statistical population/distribution of dislocations (cf. Groma 1997, 2010; Groma et al. 2003; Arsenlis and Parks 1999, 2002; El-Azab 2000, 2006; Zaiser 2001; LeSar and Rickman 2004; El-Azab et al. 2007; Hochrainer et al. 2007, 2014; Hochrainer 2015; Deng and El-Azab 2010; Ispanovity et al. 2010; Sandfeld et al. 2011; Xia and El-Azab 2015; Kapetanou et al. 2015; Groma et al. 2015, 2016; Xia et al. 2016; Monavari et al. 2016). These models employ averaging

procedures based on the statistical mechanics of interacting many-body systems, recognizing that the Nye tensor employed in FDM and numerous GC theories typically offer a description of only geometrically necessary dislocation (GND) density in an averaged sense over all the line directions in a volume element (or Burgers circuit) and not the overall dislocation population and are therefore limited in nature. Statistical descriptions of dislocation densities and fluxes are nontrivial, as pointed out by Sandfeld et al. (2011). Limitations naturally arise associated with averaging assumptions that are made. The description of dislocations goes beyond that of the classical dislocation density tensor (e.g., El-Azab 2000; Hochrainer et al. 2007) to provide more statistical information. In some cases, the total dislocation density is also tracked (i.e., statistically stored content as well as GND density) to convey more information. In its 3D implementation (cf. Xia and El-Azab 2015), distributions of the dislocation segments of edge and screw character are each tracked at each integration point, along with mobility-dependent migration with wait times at barriers. Appropriate boundary conditions are necessary as in DDD for finite domains and free surfaces.

SCD models ostensibly bridge between DDD and GC models in appearance but have strong connectivity to the governing equations of FDM, also considering statistical moments of dislocation ensembles and their interactions. Like FDM, SCD can resolve to the nanoscale, and like GC it can bridge upward to express incompatibility mechanics at scales of hundreds of nm. A principal advantage of SCD is its potential to shed light on scaling relations for dislocation substructure formation and to explore sensitivity of these relations to various physical parameters and mechanisms. Xia and El-Azab (2015) predicted arrangements of dislocations into cell structures by treating plastic deformation as an eigenstrain field, solving an incremental elastic boundary value problem for slip system driving forces and transport of dislocation densities, similar in nature to FDM models. Cross slip has been admitted at the unit process level of junction bypass (Xia and El-Azab 2015; Xia et al. 2016) and was found to play a key role in describing formation of dislocation cell structures, as was also found using DDD by Hussein and El-Awady (2016).

## 4.6    Generalized Continua (GC) Crystal Plasticity Models

Generalized continua (GC) crystal plasticity models introduce dependence on specific dislocation configurations necessary to accommodate compatible imposed strain, as reflected through scale-dependent incompatibility of the elastic and plastic deformation gradients. GC models for dislocation plasticity are necessarily of reduced order compared to the preceding frameworks. They augment strengthening due to overall dislocation density with dislocation configuration dependent strengthening expressed by dependence on the GND density, invoking a range of model assumptions in this regard. Advanced GC theories for dislocations in a lattice typically associate higher order stress quantities that conjugate with evolution of generalized kinematical variables to contribute to dissipation. They are intended to

apply to length scales well above the average spacing of dislocations and to initial-boundary value problems with scales that are typically on the order of 500 nm and above. Because they are reduced order continuum formulations and typically require introduction of normalizing characteristic length scale(s), most GC models do not offer predictive character in the same sense as DDD, FDM, and SCD models, requiring more model assumptions, phenomenology, increased number of parameters, and influence of associated calibration schemes. Still, they have value since they are highly amenable to efficient finite element implementation and can span upwards through modeling at the length scale of polycrystals. GC models are extensions of classical local crystal (LC) plasticity models that include various nonlocal approximations to incorporate effects of size and dislocation configuration. Due to space limitations and extensive coverage elsewhere, LC models (cf. Asaro 1983) are not discussed here.

GC crystal plasticity models of strain or slip gradient (Busso et al. 2000; Gurtin 2002; Evers et al. 2004; Cheong et al. 2005; Bayley et al. 2006; Gurtin and Anand 2007; Abu Al-Rub et al. 2007; Viatkina et al. 2007; Gerken and Dawson 2008; Dunne et al. 2012), micropolar (Forest et al. 2000; Forest and Sievert 2003; Mayeur and McDowell 2011, 2013, 2014, 2015) and micromorphic (Aslan et al. 2011; Cordero et al. 2010, 2013) types have advanced substantially over the past 15 years, to the point where they are viable and efficient reduced order alternatives to DDD or SCD for modeling mechanical response of domains with scales on the order of $\mu$m to tens or hundreds of $\mu$m, including polycrystals. These models employ the phenomenological notion of so-called "characteristic length scale" parameters that normalize gradient or nonlocal terms to capture size effects. These length scale parameters do not represent physical characteristic lengths but depend to varying degree on the initial conditions, boundary conditions, and level of inelastic deformation (Voyiadjis and Abu Al-Rub 2005).

Second gradient (plastic strain gradient or slip gradient) crystal plasticity models regularize the influence of strain gradients on the material workhardening behavior (Busso et al. 2000; Gurtin 2002; Evers et al. 2004; Cheong et al. 2005; Bayley et al. 2006; Gurtin and Anand 2007; Abu Al-Rub et al. 2007; Viatkina et al. 2007; Gerken and Dawson 2008; Dunne et al. 2012). Back stress arises naturally in such models as a manifestation of GND gradients and manifest specimen size and boundary effects. As a further extension, Hurtado and Ortiz (2013) have offered a novel GC crystal plasticity framework that includes both energy created by surface steps and self-energy of dislocations based on slip gradients (GNDs) that is able to capture the significant influence of free surfaces on the stress-strain behavior of micropillars in compression.

Recent developments have cast GC crystal plasticity models into forms that express sensitivity to multiphase microstructures and complex dislocation mechanisms (cf. Keshavarz and Ghosh 2013; Ghosh et al. 2016a, b; Shahba and Ghosh 2016), along with sophisticated mechanistic evolution equations at the slip system level and associated calibration schemes. It is very difficult to incorporate comparable levels of key dislocation mechanisms and material structure hierarchy in the other model constructs shown in Fig. 1 for complex alloy systems. For example,

Keshavarz and Ghosh (2015) proposed a hierarchical GC crystal plasticity for Ni-base superalloys. Ghosh et al. (2016b) introduced a parametrically homogenized GC crystal plasticity model for γ-γ′ Ni-base superalloys that addresses transitions across three scales: subgrain precipitates and dislocation-precipitate interactions, grain level response, and collective polycrystal behavior with realistic 3D microstructures.

The extension of GC crystal plasticity to model submicron scale processes involving dislocation substructure and slip transfer at grain boundaries is still in early stages (Bieler et al. 2009). Microstructure-sensitive GC models that incorporate evolving dislocation density at the slip system level as internal variables (cf. Raabe et al. 2001; Ma et al. 2006; Roters et al. 2010; Dunne et al. 2012; Keshavarz and Ghosh 2013; Mayeur and McDowell 2013, 2014, 2015; Shahba and Ghosh 2016) offer enhanced correlative capabilities for complex loading conditions in many cases at the scale of individual grains. Still, the experimentally observed number of activated slip systems at the subgrain scale is typically limited compared to simulations using most mesoscopic crystal plasticity models of GC and LC type (Lloyd 2010; Buchheit et al. 2005); this is an issue that requires increased focus on slip system hardening relations and more intimate connection to in situ experimental data such as EBSD and digital image correlation measurements of surface deformation.

## 5   Summary and Challenges

Some of the more challenging and longstanding mesoscale gaps for the hierarchy of crystalline plasticity models listed in Fig. 1 include:

- Distinguishing dislocation nucleation and growth processes.
- Initial conditions and evolving structure of interfaces.
- Slip transfer at interfaces, including interface evolution and damage.
- Quantifying coarse-graining and model reduction errors.
- Uncertainty quantification of individual models, domain decomposition, and various two-scale transitions.
- Modeling dislocations in multicomponent material systems (cf. Rao et al. 2017), including appropriate interatomic potentials for atomistic methods that include uncertainty quantification.
- Improving fidelity and accuracy of simulation of local states of slip system activation and degree of slip within individual grains to support material fatigue, fracture, and ductility assessments.
- Coupling dislocation structure and behavior with point defect interactions, including climb, strain aging, and hydrogen, interstitial, and vacancy interactions.
- Timescaling across a spectrum of thermally activated processes involving a range of activation volumes.

    In spite of significant limitations in timescaling and incorporation of finite temperature effects, atomistic coarse-graining methods have more predictive

character than other scale transition schemes to support understanding and design of new material systems and microstructures affected by dislocation phenomena. DDD and SCD models hold great promise to predict scale-dependent dislocation strengthening, but at present are subject to greater uncertainty in addressing realistic dislocation microstructure effects such as grain or phase boundaries, crystal anisotropy, etc. at substantial strain levels. Higher length scale models (e.g., GC or LC) and associated scale transition schemes are appropriate more for systems having rather well-established understanding of dislocation mechanisms. The concern for GC and LC crystal plasticity models is their lack of predictive support for discovery and design of new alloy systems that do not have an established knowledge base. To some extent, DDD also relies on the prior knowledge base for materials (witness the large number of applications to pure metals in the literature), as well as uncertain information from atomistic models to inform parameters and algorithmic choices.

# References

Abu Al-Rub RK, Voyiadjis GZ, Bammann DJ (2007) A thermodynamic based higher-order gradient theory for size dependent plasticity. Int J Solids Struct 44:2888–2923

Acharya A (2001) A model of crystal plasticity based on the theory of continuously distributed dislocations. J Mech Phys Solids 49:761–784

Acharya A (2003) Driving forces and boundary conditions in continuum dislocation mechanics. Proc R Soc Lond A 459:1343–1363

Acharya A, Roy A, Sawant A (2006) Continuum theory and methods for coarse-grained, mesoscopic plasticity. Scr Mater 54:705–710

Acharya A, Beaudoin AJ, Miller R (2008) New perspectives in plasticity theory: dislocation nucleation, waves and partial continuity of the plastic strain rate. Math Mech Solids 13(3–4):292–315

Akarapu S, Zbib HM, Bahr DF (2010) Analysis of heterogeneous deformation and dislocation dynamics in single crystal micropillars under compression. Int J Plast 16:239–257

Amelang JS, Venturini GN, Kochmann DM (2013) Microstructure evolution during nanoindentation by the quasicontinuum method. Proc Appl Math Mech 13:553–556

Amelang JS, Venturini GN, Kochmann DM (2015) Summation rules for a fully nonlocal energy-based quasicontinuum method. J Mech Phys Solids 82:378–413

Amodeo RJ, Ghoniem NM (1988) A review of experimental-observations and theoretical-models of dislocation cells and subgrains. Res Mechanica 23(2–3):137–160

Amodeo RJ, Ghoniem NM (1990) Dislocation dynamics. I. A proposed methodology for micromechanics. Phys Rev B 41:6958

Arsenlis A, Parks DM (1999) Crystallographic aspects of geometrically-necessary and statistically-stored dislocation density. Acta Mater 47(5):1597–1611

Arsenlis A, Parks DM (2002) Modeling the evolution of crystallographic dislocation density in crystal plasticity. J Mech Phys Solids 50:1979–2009

Arsenlis A, Cai W, Tang M, Rhee M, Oppelstrup T, Hommes G, Pierce TG, Bulatov VV (2007) Enabling strain hardening simulations with dislocation dynamics. Model Simul Mater Sci Eng 15:553–595

Asaro RJ (1983) Crystal plasticity. ASME J Appl Mech 50:921–934

Aslan O, Cordero NM, Gaubert A, Forest S (2011) Micromorphic approach to single crystal plasticity and damage. Int J Eng Sci 49:1311–1325

Baker KL, Curtin WA (2016) Multiscale diffusion method for simulations of long-time defect evolution with application to dislocation climb. J Mech Phys Solids 92: 297–312

Bayley CJ, Brekelmans WAM, Geers MGD (2006) A comparison of dislocation-induced back stress formulations in strain gradient crystal plasticity. Int J Solids Struct 43:7268–7286

Bertin N, Upadhyay MV, Pradalier C, Capolungo L (2015) A FFT-based formulation for efficient mechanical fields computation in isotropic and anisotropic periodic discrete dislocation dynamics. Model Simul Mater Sci Eng 23(6):065009

Bieler TR, Eisenlohr P, Roters F, Kumar D, Mason DE, Crimp MA, Raabe D (2009) The role of heterogeneous deformation on damage nucleation at grain boundaries in single phase metals. Int J Plast 25:1655–1683

Binder A, Luskin M, Perez D, Voter AF (2015) Analysis of transition state theory rates upon spatial coarse-graining. Multiscale Model Simul 13:890–915

Buehler MJ, Hartmaier A, Duchaineau MA, Abraham FF, Gao H (2005) The dynamical complexity of work-hardening: a large-scale molecular dynamics simulation. Acta Mech Sinica 21: 103–111

Buchheit TE, Wellman GW, Battaille C (2005) Investigating the limits of polycrystal plasticity modeling. Int J Plast 21(2):221–249

Busso EP, Meissonnier FT, O'Dowd NP (2000) Gradient-dependent deformation of two-phase single crystals. J Mech Phys Solids 48:2333–2361

Cai W, Arsenlis A, Weingberger CR, Bulatov VV (2006) A non-singular continuum theory of dislocations. J Mech Phys Solids 54:561–587

Cai W, Sills RB, Barnett DM, Nix WD (2014) Modeling a distribution of point defects as misfitting inclusions in stressed solids. J Mech Phys Solids 66:154–171

Chen L, Chen J, Lebensohn R, Chen L-Q (2014) An integrated fast Fourier transform-based phase-field and crystal plasticity approach to model recrystallization of three dimensional polycrystals. Comput Methods Appl Mech Eng 285:829–848

Chen Y (2009) Reformulation of microscopic balance equations for multiscale materials modeling. J Chem Phys 130:134706

Cheong KS, Busso EP, Arsenlis A (2005) A study of microstructural length scale effects on the behaviour of fcc polycrystals using strain gradient concepts. Int J Plast 21:1797–1814

Cho J, Junge T, Molinari F-F, Anciaux G (2015) Toward a 3D coupled atomistics and discrete dislocation dynamics simulation: dislocation core structures and Peierls stresses with several character angles in FCC aluminum. Adv Model Simul Eng Sci 2:12. https://doi.org/10.1186/s40323-015-0028-6

Cordero NM, Gaubert A, Forest S, Busso EP, Gallerneau F, Kruch S (2010) Size effects in generalised continuum crystal plasticity for two-phase laminates. J Mech Phys Solids 58: 1963–1994

Cordero NM, Forest S, Busso EP (2013) Micromorphic modelling of grain size effects in metal polycrystals. GAMM-Mitteilungen 36(2):186–202

Crone JC, Chung PW, Leiter KW, Knap J, Aubry S, Hommes G, Arsenlis A (2014) A multiply parallel implementation of finite element-based discrete dislocation dynamics for arbitrary geometries. Model Simul Mater Sci Eng 22:035014–035041

Deng J, El-Azab A (2010) Temporal statistics and coarse graining of dislocation ensembles. Philos Mag 90(27–28):3651–3678

Devincre, B, Madec R, Monnet G, Queyreau S, Gatti R, Kubin L (2011) Modeling crystal plasticity with dislocation dynamics simulations: the 'microMegas' code. In: Mechanics of nano-objects. Presses del'Ecole des Mines de Paris, Paris, pp 81–100

Dewald MP, Curtin WA (2007a) Multiscale modelling of dislocation/grain boundary interactions. II. Screw dislocations impinging on tilt boundaries in Al. Philos Mag 87:4615–4641

Dewald MP, Curtin WA (2007b) Multiscale modeling of dislocation/grain boundary interactions: I. Edge dislocations impinging on S11 (113) tilt boundary in Al. Model Simul Mater Sci Eng 15:S193–S215

Dewald MP, Curtin WA (2011) Multiscale modeling of dislocation/grain-boundary interactions: III. 60° dislocations impinging on Σ3:Σ9 and Σ11 tilt boundaries in Al. Model Simul Mater Sci Eng 19:055002

Dunne FPE, Kiwanuka R, Wilkinson AJ (2012) Crystal plasticity analysis of micro-deformation, lattice rotation and geometrically necessary dislocation density. Proc R Soc A 468:2509–2531

Eidel B, Stukowski A (2009) A variational formulation of the quasicontinuum method based on energy sampling in clusters. J Mech Phys Solids 57:87–108

El-Azab A (2000) Statistical mechanics treatment of the evolution of dislocation distributions in single crystals. Phys Rev B 61(18):11956–11966

El-Azab A (2006) Statistical mechanics of dislocation systems. Scr Mater 54:723–727

El-Azab A, Deng J, Tang M (2007) Statistical characterization of dislocation ensembles. Philos Mag 87(8–9):1201–1223

Evers LP, Brekelmans WAM, Geers MGD (2004) Non-local crystal plasticity model with intrinsic SSC and GND effects. J Mech Phys Solids 52:2379–2401

Forest S, Sievert R (2003) Elastoviscoplastic constitutive frameworks for generalized continua. Acta Mech 160:71–111

Forest S, Barbe F, Cailletaud G (2000) Cosserat modelling of size effects in the mechanical behaviour of polycrystals and multi-phase materials. Int J Solids Struct 37:7105–7126

Gerken JM, Dawson PR (2008) A crystal plasticity model that incorporates stresses and strains due to slip gradients. J Mech Phys Solids 56(4):1651–1672

Ghoniem NM (2005) A perspective on dislocation dynamics. In: Yip S (ed) Handbook of materials modeling, vol 1. Methods and models. Springer, Dordrecht, pp 1–7

Ghosh S, Lee K, Raghavan P (2001) A multi-level computational model for multi-scale damage analysis in composite and porous materials. Int J Solids Struct 38(14):2335–2385

Ghosh S, Bai J, Raghavan P (2007) Concurrent multi-level model for damage evolution in micro structurally debonding composites. Mech Mater 39(3):241–266

Ghosh S, Shahba A, Tu X, Huskins EL, Schuster BE (2016a) Crystal plasticity FE modeling of Ti alloys for a range of strain-rates. Part II: image-based model with experimental validation. Int J Plast 87:69–85

Ghosh S, Weber G, Keshavarz S (2016b) Multiscale modeling of polycrystalline nickel-based superalloys accounting for subgrain microstructures. Mech Res Commun 78:34–46

Groh S, Zbib HM (2009) Advances in discrete dislocations dynamics and multiscale modeling. J Eng Mater Technol 131:041209-1–041209-10

Groh S, Marin EB, Horstemeyer MF, Zbib HM (2009) Multiscale modeling of the plasticity in an aluminum single crystal. Int J Plast 25:1456–1473

Groma I (1997) Link between the microscopic and mesoscopic length-scale description of the collective behavior of dislocations. Phys Rev B 56(10):5807–5813

Groma I (2010) Statistical physical approach to describe the collective properties of dislocations. In: Gumbsch P, Pippan R (eds) Multiscale modelling of plasticity and fracture by means of dislocation mechanics. CISM International Centre for Mechanical Sciences, Vienna, pp 213–270. ISBN 978-3-7091-0283-1

Groma I, Csikor FF, Zaiser M (2003) Spatial correlations and higher-order gradient terms in a continuum description of dislocation dynamics. Acta Mater 51:1271

Groma I, Vandrus Z, Ispanovity PD (2015) Scale-free phase field theory of dislocations. Phys Rev Lett 114:015503

Groma I, Zaiser M, Ispanovity PD (2016) Dislocation patterning in a 2D continuum theory of dislocations. arXiv:1601.07831 [cond-mat.mtrl-sci]

Gulluoglu AN, Srolovitz DJ, Lesar R, Lomdahl PS (1989) Dislocation distributions in two dimensions. Scr Metall 23:1347–1352

Gurtin ME (2002) A gradient theory of single-crystal viscoplasticity that accounts for geometri-
cally necessary dislocations. J Mech Phys Solids 50(1):5–32

Gurtin ME, Anand L (2007) A gradient theory for single-crystal plasticity. Model Simul Mater Sci
Eng 15:S263–S270

Guruprasad PJ, Benzerga AA (2008) Size effects under homogeneous deformation of single
crystals: a discrete dislocation analysis. J Mech Phys Solids 56:132–156

Hartley CS, Mishin Y (2005) Representation of dislocation cores using Nye tensor distributions.
Mater Sci Eng A 400–401:18–21

Hochrainer T (2015) Multipole expansion of continuum dislocation dynamics in terms of
alignment tensors. Philos Mag 95:1321–1367

Hochrainer T, Zaiser M, Gumbsch P (2007) A three-dimensional continuum theory of dislocation
systems: kinematics and mean-field formulation. Philos Mag 87:1261–1282

Hochrainer T, Sandfeld S, Zaiser M, Gumbsch P (2014) Continuum dislocation dynamics: towards
a physical theory of crystal plasticity. J Mech Phys Solids 63:167–178

Hurtado DE, Ortiz M (2013) Finite element analysis of geometrically necessary dislocations in
crystal plasticity. Int J Numer Methods Eng 93:66–79

Hussein AM, El-Awady JA (2016) Quantifying dislocation microstructure evolution and cyclic
hardening in fatigued face-centered cubic single crystals. J Mech Phys Solids 91:126–144

Irving J, Kirkwood J (1950) The statistical mechanical theory of transport processes. IV The
equations of hydrodynamics. J Chem Phys 8:817–829

Kang K, Yin J, Cai W (2014) Stress dependence of cross slip energy barrier for face-centered cubic
nickel. J Mech Phys Solids 62:181–193

Kapetanou O, Koutsos V, Theotokoglou E, Weygand D, Zaiser M (2015) Statistical analysis
and stochastic dislocation based modeling of microplasticity. J Mech Behav Mater 24(3–4):
105–113

Keralavarma SM, Benzerga AA (2015) High-temperature discrete dislocation plasticity. J Mech
Phys Solids 82:1–22

Keshavarz S, Ghosh S (2013) Multi-scale crystal plasticity finite element model approach to
modeling nickel-based superalloys. Acta Mater 61:6549–6561

Keshavarz S, Ghosh S (2015) Hierarchical crystal plasticity FE model for nickel-based superalloys:
sub-grain microstructures to polycrystalline aggregates. Int J Solids Struct 55:17–31

Khraishi T, Zbib HM (2002) Free-surface effects in 3D dislocation dynamics: formulation and
modeling. ASME J Eng Mater Technol 124(3):342–351

Kim WK, Luskin M, Perez D, Voter AF, Tadmor EB (2014) Hyper-QC: an accelerated finite-
temperature quasicontinuum method using hyperdynamics. J Mech Phys Solids 63:94–112

Kirkwood JG (1946) The statistical mechanical theory of transport processes. I. General theory. J
Chem Phys 14:180

Knap J, Ortiz M (2001) An analysis of the quasicontinuum method. J Mech Phys Solids 49:
1899–1923

Kochmann DM, Venturini GN (2014) A meshless quasicontinuum method based on local
maximum-entropy interpolation. Model Simul Mater Sci Eng 22:034007–034035

Kouznetsova VG, Geers MGD, Brekelmans WAM (2004) Multi-scale second-order computational
homogenization of multi-phase materials: a nested finite element solution strategy. Comput
Methods Appl Mech Eng 193(48–51):5525–5550

Kwon S, Lee Y, Park JY, Sohn D, Lim JH, Im S (2009) An efficient three-dimensional adaptive
quasicontinuum method using variable-node elements. J Comput Phys 228:4789–4810

Lepinoux J, Kubin LP (1987) The dynamic organization of dislocation structures – a simulation.
Scr Metall 21:833–838

LeSar R, Rickman JM (2004) Incorporation of local structure in continuous theory of dislocations.
Phys Rev B 69:172105

Liu B, Arsenlis A, Aubry S (2016) Computing forces on interface elements exerted by dislocations
in an elastically anisotropic crystalline material. Model Simul Mater Sci Eng 24:055013

Lloyd JT (2010) Implications of limited slip in crystal plasticity. M.S. Thesis, Woodruff School of
Mechanical Engineering, Georgia Institute of Technology

Ma A, Roters F, Raabe D (2006) A dislocation density based constitutive model for crystal plasticity FEM including geometrically necessary dislocations. Acta Mater 54:2169–2179

Marian J, Venturini G, Hansen BL, Knap J, Ortiz M, Campbell GH (2010) Finite-temperature extension of the quasicontinuum method using Langevin dynamics: entropy losses and analysis of errors. Model Simul Mater Sci Eng 18(1):015003

Martınez E, Mariana J, Arsenlis A, Victoria M, Perlado JM (2008) Atomistically informed dislocation dynamics in fcc crystals. J Mech Phys Solids 56:869–895

Matveev MV, Selivanikova OV, Cherepanov DN (2016) Formation of deformation substructures in FCC crystals under the influence of point defect fluxes. Mater Sci Eng 124:012129

Mayeur JR, McDowell DL (2011) Bending of single crystal thin films as predicted by micropolar crystal plasticity. Int J Eng Sci 49:1357–1366

Mayeur JR, McDowell DL (2013) An evaluation of higher-order single crystal strength models for constrained thin films subjected to simple shear. J Mech Phys Solids 61(9):1935–1954

Mayeur JR, McDowell DL (2014) A comparison of Gurtin-type and micropolar single crystal plasticity with generalized stresses. Int J Plast 57:29–51

Mayeur JR, McDowell DL (2015) Micropolar crystal plasticity simulations of particle strengthening. Model Simul Mater Sci Eng 23(6):065007

Mayeur JR, McDowell DL, Bammann DJ (2011) Dislocation-based micropolar single crystal plasticity: comparison of multi- and single-criterion theories. J Mech Phys Solids 59(2): 398–422

McDowell DL (2008) Viscoplasticity of heterogeneous metallic materials. Mater Sci Eng R: Rep 62(3):67–123

McDowell DL (2010) A perspective on trends in multiscale plasticity. Int J Plast 26(9):1280–1309

Mendis BG, Mishin Y, Hartley CS, Hemker KJ (2006) Use of the Nye tensor in analyzing HREM images of bcc screw dislocations. Philos Mag 86(29–31):4607–4640

Miller R, Tadmor EB, Phillips R, Ortiz M (1998a) Quasicontinuum simulation of fracture at the atomic scale. Model Simul Mater Sci Eng 6(5):607–638

Miller R, Ortiz M, Phillips R, Shenoy V, Tadmor EB (1998b) Quasicontinuum models of fracture and plasticity. Eng Fract Mech 61(3–4):427–444

Monavari M, Sandfeld S, Zaiser M (2016) Continuum representation of systems of dislocation lines: a general method for deriving closed-form evolution equations. J Mech Phys Solids 95:575–601

Nguyen LD, Baker KL, Warner DH (2011) Atomistic predictions of dislocation nucleation with transition state theory. Phys Rev B 84:024118

Nye JF (1953) Some geometrical relations in dislocated crystals. Acta Metall 1:153–162

Pavia F, Curtin WA (2015) Parallel algorithm for multiscale atomistic/continuum simulations using LAMMPS. Model Simul Mater Sci Eng 23:055002 (23 pp)

Peierls R (1940) The size of a dislocation. Proc Phys Soc Lond 52:34–37

Pluchino PA, Chen X, Garcia M, Xiong L, McDowell DL, Chen Y (2016) Dislocation migration across coherent phase interfaces in SiGe superlattices. Comput Mater Sci 111:1–6

Qu S, Shastry V, Curtin WA, Miller RE (2005) A finite-temperature dynamic coupled atomistic/discrete dislocation method. Model Simul Mater Sci Eng 13(7):1101–1118

Raabe D, Sachtleber M, Zhao Z, Roters F, Zaefferer S (2001) Micromechanical and macromechanical effects in grain scale polycrystal plasticity experimentation and simulation. Acta Mater 49(17):3433–3441

Rao SI, Woodward C, Parthasarathy TA, Senkov O (2017) Atomistic simulations of dislocation behavior in a model FCC multicomponent concentrated solid solution alloy. Acta Mater 134:188–194

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58(4):1152–1211

Roy A, Acharya A (2006) Size effects and idealized dislocation microstructure at small scales: predictions of a phenomenological model of mesoscopic field dislocation mechanics. J Mech Phys Solids 54:1711–1743

Roy A, Puri S, Acharya A (2007) Phenomenological mesoscopic dislocation mechanics, lower-order gradient plasticity and transport of mean excess dislocation density. Model Simul Mater Sci Eng 15:S167–S180

Rudd RE, Arsenlis A, Barton NR, Cavallo RM, Comley AJ, Maddox BR, Marian J, Park H-S, Prisbrey ST, Wehrenberg CE, Zepeda-Ruiz L, Remington BA (2014) Multiscale strength (MS) models: their foundation, their successes, and their challenges. In: 18th APS-SCCM and 24th AIRAPT Journal of Physics: Conference Series 500, p 112055

Ryu S, Kang K, Cai W (2011) Predicting the dislocation nucleation rate as a function of temperature and stress. J Mater Res 26(18):2335–2354

Ryu I, Nix WD, Cai W (2013) Plasticity of bcc micropillars controlled by competition between dislocation multiplication and depletion. Acta Mater 61:3233–3241

Sandfeld S, Hochrainer T, Zaiser M, Gumbsch P (2011) Continuum modeling of dislocation plasticity: theory, numerical implementation and validation by discrete dislocation simulations. J Mater Res 26:623–632

Saroukhani S, Nguyen LD, Leung KWK, Singh CV, Warner DH (2016) Harnessing atomistic simulations to predict the rate at which dislocations overcome obstacles. J Mech Phys Solids 90:203–214

Shahba A, Ghosh S (2016) Crystal plasticity FE modeling of Ti alloys for a range of strain-rates. Part I: a unified constitutive model and flow rule. Int J Plast 87:48–68

Shen C, Wang Y (2003) Modeling dislocation network and dislocation–precipitate interaction at mesoscopic scale using phase field method. Int J Multiscale Comput Eng 1(1): 91–104

Shen C, Li J, Wang Y (2014) Predicting structure and energy of dislocations and grain boundaries. Acta Mater 74:125–131

Shenoy VB, Miller R, Tadmor EB, Phillips R, Ortiz M (1998) Quasicontinuum models of interfacial structure and deformation. Phys Rev Lett 80(4):742–745

Shenoy VB, Miller R, Tadmor E, Rodney D, Phillips R, Ortiz M (1999) An adaptive finite element approach to atomic-scale mechanics - the quasicontinuum method. J Mech Phys Solids 47(3):611–642

Shiari B, Miller RE, Curtin WA (2005) Coupled atomistic/discrete dislocation simulations of nanoindentation at finite temperature. ASME J Eng Mater Technol 127(4):358–368

Shilkrot LE, Curtin WA, Miller RE (2002a) A coupled atomistic/continuum model of defects in solids. J Mech Phys Solids 50:2085–2106

Shilkrot LE, Miller RE, Curtin WA (2002b) Coupled atomistic and discrete dislocation plasticity. Phys Rev Lett 89:025501–025501

Shilkrot LE, Miller RE, Curtin WA (2004) Multiscale plasticity modeling: coupled atomistics and discrete dislocation mechanics. J Mech Phys Solids 52:755–787

Shimokawa T, Kinari T, Shintaku S (2007) Dislocation-grain boundary interactions by the Quasicontinuum method. Key Eng Mater 340–341:973–978

Sills RB, Aghaei A, Cai W (2016) Advanced time integration algorithms for dislocation dynamics simulations of work hardening. Model Simul Mater Sci Eng 24:045019 (17 pp)

Tadmor EB, Miller RE (2011) Modeling materials: continuum, atomistic and multiscale techniques. Cambridge University Press, New York

Tadmor EB, Ortiz M, Phillips R (1996a) Quasicontinuum analysis of defects in solids. Philos Mag A 73(6):1529–1563

Tadmor EB, Phillips R, Ortiz M (1996b) Mixed atomistic and continuum models of deformation in solids. Langmuir 12(19):4529–4534

Tang M, Hommes G, Aubry S, Arsenlis A (2011) ParaDiS-FEM dislocation dynamics simulation code primer. LLNL-TR-501662. https://doi.org/10.2172/1037843

Tembhekar I, Amelang JS, Munk L, Kochmann DM (2017) Automatic adaptivity in the fully non-local quasicontinuum method for coarse-grained atomistic simulations. Int J Numer Methods Eng 110:878–900

Van der Giessen E, Needleman A (1995) Discrete dislocation plasticity: a simple planar model. Model Simul Mater Sci Eng 3:689–735

Viatkina EM, Brekelmans WAM, Geers MGD (2007) Modelling of the internal stress in dislocation cell structures. Eur J Mech A Solids 26:982–998

Voyiadjis GZ, Abu Al-Rub RK (2005) Gradient plasticity theory with a variable length scale parameter. Int J Solids Struct 42(14):3998–4029

Wallin M, Curtin WA, Ristinmaa M, Needleman A (2008) Multi-scale plasticity modeling: coupled discrete dislocation and continuum crystal plasticity. J Mech Phys Solids 56:3167–3180

Wang Y, Li J (2010) Phase field modeling of defects and deformation. Acta Mater 58(4):1212–1235

Wang YU, Jin YM, Cuitino AM, Khachaturyan AG (2001) Nanoscale phase field microelasticity theory of dislocations: model and 3D simulations. Acta Mater 49(10):1847–1857

Wang W, Ghoniem N, Swaminaryan LSR (2006) A parallel algorithm for 3D dislocation dynamics. J Comput Phys 219:608–621

Xia S, El-Azab A (2015) Computational modelling of mesoscale dislocation patterning and plastic deformation of single crystals. Model Simul Mater Sci Eng 23(5):55009–55034

Xia S, Belak J, El-Azab A (2016) The discrete-continuum connection in dislocation dynamics: I. Time coarse graining of cross slip. Modell Simul Mater Sci Eng 24:075007 (22pp)

Xiong L, Tucker GJ, McDowell DL, Chen Y (2011) Coarse-grained atomistic simulation of dislocations. J Mech Phys Solids 59:160–177

Xiong L, Deng Q, Tucker GJ, McDowell DL, Chen Y (2012a) A concurrent scheme for passing dislocations from atomistic to continuum regions. Acta Mater 60(3):899–913

Xiong L, Deng Q, Tucker GJ, McDowell DL, Chen Y (2012b) Coarse-grained atomistic simulations of dislocations in Al, Ni and Cu crystals. Int J Plast 38:86–101

Xiong L, McDowell DL, Chen Y (2014a) Sub-THz phonon drag on dislocations by coarse-grained atomistic simulations. Int J Plast 55:268–278

Xiong L, Chen X, McDowell DL, Chen Y (2014b) Predicting phonon properties of 1D polyatomic crystals through the concurrent atomistic-continuum simulations. Arch Appl Mech (special issue in honor of Professor G. Maugin) 84:1665–1675

Xiong L, Xu S, McDowell DL, Chen Y (2015) Concurrent atomistic-continuum simulations of dislocation-void interactions in fcc crystals. Int J Plast 65:33–42

Xu S, Che R, Xiong L, Chen Y, McDowell DL (2015) A quasistatic implementation of the concurrent atomistic-continuum method for FCC crystals. Int J Plast 72:91–126

Xu S, Xiong L, Chen Y, McDowell DL (2016) Sequential slip transfer of mixed character dislocations across Σ3 coherent twin boundary in FCC metals: a concurrent atomistic-continuum study. npg Comput Mater 2:15016. https://doi.org/10.1038/npjcompumats.2015.16

Xu S, Xiong L, Chen Y, McDowell DL (2017) Comparing EAM potentials to model slip transfer of sequential mixed character dislocations across two symmetric tilt grain boundaries in Ni. JOM 69(5):814–821

Yasin H, Zbib HM, Khaleel MA (2001) Size and boundary effects in discrete dislocation dynamics: coupling with continuum finite element. Mater Sci Eng A 309–310:294–299

Yin J, Barnett DM, Fitzgerald SP, Cai W (2012) Computing dislocation stress fields in anisotropic elastic media using fast multipole expansions. Model Simul Mater Sci Eng 20:045015

Yu W, Wang Z (2012) Interactions between edge lattice dislocations and Σ11 symmetrical tilt grain boundaries in copper: a quasi-continuum method study. Acta Mater 60(13–14):5010–5021

Yu WS, Wang ZQ (2014) Interactions between edge lattice dislocations and sigma 11 symmetrical tilt grain boundary: comparisons among several FCC metals and interatomic potentials. Philos Mag 94:2224–2246

Zaiser M (2001) Statistical modeling of dislocation systems. Mater Sci Eng A 309-310:304–315

Zbib HM, Diaz de la Rubia T (2002) A multiscale model of plasticity. Int J Plast 18(9):1133–1163

Zbib HM, Rhee M, Hirth JP (1998) On plastic deformation and the dynamics of 3D dislocations. Int J Mech Sci 40(2):113–127

Zbib HM, Diaz de la Rubia T, Bulatov V (2002) A multiscale model of plasticity based on discrete dislocation dynamics. ASME J Eng Mater Technol 124(1):78–87

Zbib HM, Overman CT, Akasheh F, Bahr D (2011) Analysis of plastic deformation in nanoscale metallic multilayers with coherent and incoherent interfaces. Int J Plast 27:1618–1639

Zhang X (2015) Field dislocation mechanics with applications in atomic, mesoscopic and tectonic scale problems. Dissertations, Carnegie Mellon University Paper 585. http://repository.cmu.edu/cgi/viewcontent.cgi?article=1624&context=dissertations

# Developing Virtual Microstructures and Statistically Equivalent Representative Volume Elements for Polycrystalline Materials

**76**

Somnath Ghosh and Michael A. Groeber

## Contents

**Abstract**

This chapter introduces computational methods for generating virtual material microstructures of engineering materials with heterogeneities. Microstructures of polycrystalline materials containing localized features such as annealing

S. Ghosh (✉)
JHU Center for Integrated Structure-Materials Modeling and Simulation (CISMMS), Air Force Center of Excellence on Integrated Materials Modeling (CEIMM), US Association of Computational Mechanics (USACM), Department of Civil Engineering, Departments of Mechanical Engineering and Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA
e-mail: sghosh20@jhu.edu

M. A. Groeber
Wright Patterson Air Force Base, Air Force Research Laboratory, Dayton, OH, USA
e-mail: michael.groeber@us.af.mil

twins, particulates or precipitates, and subgrain phases are the focus of this discussion. The methods use data from characterization methods to provide 3D statistical distribution and correlation functions that serve as inputs to the virtual microstructure generation process. Computational methods infer 3D statistical descriptors from 2D surface data and use stereology or other optimization-based projection techniques for 2D to 3D development. The chapter reviews the DREAM.3D software package and discusses newly developed methods to incorporate twins, particles, and subgrain-scale phases. Finally, the microstructure-based SERVE is introduced in the realm of establishing microstructure-property relations.

# 1    Introduction

The recent years have seen an increasing trend in the use of microstructure-based mechanistic models for predicting material deformation and extreme behavior like fracture and fatigue response. A primal need in the development of image-based micromechanical models is the creation of representative 3D virtual models of the microstructures. An obvious path is by direct image reconstruction from 3D experimental data on sample volumes of the microstructure. Such experimental data may be extracted from electron backscatter diffraction (EBSD) or scanning electron microscopy (SEM) images of serial-sectioning samples (Groeber et al. 2006) or from various computed tomography techniques (Turner et al. 2017). Deterministic models of the microstructure, representing the exact data from experiments (Bhandari et al. 2007), however, are not necessarily best suited for micromechanical simulations, since the microstructure itself may have significant spatial variations.

A less direct but effective approach is to represent structure through the use of tools that generate statistical distribution functions equivalent to desired sets obtained from experimental observations. These "statistically equivalent" virtual microstructures must capture the statistics of characteristic variables, such as grain shape and size, crystallographic orientations, and misorientations and their correlations (Groeber et al. 2008a, b). The approach generally entails quantification of experimental data followed by microstructure generation that statistically match material measurements to a predetermined degree of accuracy. It is capable of limiting the need for abundant data collection, as well as supplementing information when direct 3D data is unavailable. Furthermore it enables the incorporation of microstructural statistics in higher length-scale constitutive relations for microstructure-property relations. Finally, the ability to generate virtual microstructural instantiations allows for virtual design or sampling of the potential microstructural space, driving toward tailoring materials structure.

Creating statistically "accurate" material instantiations for many engineering materials is still in its infancy with many gaps and opportunities. Recent efforts, centered around integrated computational materials engineering (ICME) and the materials genome initiative (MGI), assume at their core that both the material

structure and functionality of materials can be captured in a digital environment. It is imperative to the success of these initiatives that the materials community develop methodologies for creating digital analogues to real materials. In practice, it is common to generate performance models with limited amount of microstructural information. Typically, simple geometric shapes or tessellations are used to represent microstructures with consideration of lower moments such as average values. The use of such lower moments is often inadequate for accurate prediction of many properties of interest, such as fatigue, fracture, creep, etc. Higher statistical moments of the microstructural distributions, depicting extreme values, are needed for these predictions. While 3D data collection plays a critical role, especially when extreme properties are of interest, 3D experimental methods alone are often not adequate for this framework. Many microstructural arrangements must be interrogated for probabilistic understanding of the relation between microstructure and these properties. To facilitate this, experimental data should be coupled with characterization methods to provide 3D statistical distribution and correlation functions that serve as the inputs to the synthetic microstructure generation process. An additional incentive in this development is to infer 3D statistical descriptors from 2D observations and surface data through the use of stereology or other projection techniques. This is particularly relevant due to the fact that 3D experimental techniques remain unavailable or prohibitively expensive to a large portion of the materials community.

Various methods have been proposed in the literature for generating polycrystalline or polyphase microstructures, e.g., in Sundararaghavan and Zabaras (2005), Kumar et al. (2016), Rollett et al. (2007), Saylor et al. (2004), Guo et al. (2014), Jiao et al. (2007, 2013) and Hasanabadi et al. (2016). In Sundararaghavan and Zabaras (2005), reconstruction of 3D microstructures is solved as a pattern recognition problem, where a microstructure database is used with limited statistical information available from planar images. Microstructures are represented in the form of undirected probabilistic graphs or Markov random fields for computing probability distribution of statistically similar microstructures in Kumar et al. (2016). In Rollett et al. (2007), a 3D grain-structure generation method is based on statistical data gathered from sections on different planes with assigned orientations. Statistically representative polycrystalline microstructures are computationally simulated in Saylor et al. (2004) from geometric and crystallographic observations from orthogonal sections. Maps on the orthogonal planes characterize the sizes, shapes, and orientations of grains, and a voxel-based tessellation technique is subsequently used to generate the microstructure. A dilation-erosion method is developed in Guo et al. (2014) for stochastic reconstruction of 3D duplex stainless steel microstructure containing percolating filamentary ferrite phase from 2D optical micrographs. In Jiao et al. (2007), the authors have concluded that the two-point correlation function space of a statistically homogeneous material can be expressed through a map, constructed on a selected set of bases of the function space. A procedure to model and predict microstructure evolution of lead-tin alloys has been developed using the two-point correlation function associated with different phases in Jiao et al. (2013). A method for 3D microstructure reconstruction from two-point correlation

functions of 2D cross sections using conditional probability theorems and a phase-recovery algorithm is developed in Hasanabadi et al. (2016). The *n*-point correlation functions have been further used in Tewari et al. (2004) and Niezgoda et al. (2010) to reconstruct the microstructure and obtain homogenized properties. Recent work by the author's group combines optimization tools like Genetic Algorithms (Goldberg 1989) with stereology-based projection techniques to develop 3D microstructures from 2D data (Pinz et al. 2018; Tu et al. 2019).

Direct numerical simulations (DNS) of large microstructural regions can be computationally prohibitive. To negotiate this, representative computational domains must be optimally defined for evaluating effective properties without having to solve large microstructural regions. This has led to the concept of a representative volume element (RVE), which is foundational to computational estimates of structure-property relations. Originally introduced in Hill (1963) as a microstructural sub-domain that is representative of the entire microstructure in an average sense, the RVE definition has undergone variations (Ostoja-Starzewski 2006; Torquato 2002; Pyrz 2006). For microstructures with nonuniformly dispersed heterogeneities, the statistically equivalent RVE or SERVE has been defined in Swaminathan et al. (2006), Swaminathan and Ghosh (2006) and McDowell et al. (2011) as the smallest microstructural domain, for which statistical distribution functions of morphological parameters, as well as material properties, converge to those for the entire microstructure. Based on the convergence property in focus, the SERVE can be classified into two categories, viz., (i) the microstructure-based SERVE or *M-SERVE*, in which morphological and crystallographic characteristics of the microstructure are the sole determinants of the representative volume, and (ii) the property-based SERVE or *P-SERVE* that are determined from convergence of selected material properties.

This chapter is aimed at discussing methods of generating 3D statistically equivalent virtual microstructures and M-SERVEs of structural materials that are characterized by polycrystalline and/or polyphase microstructures. It begins with a description of the open-source software package DREAM.3D, which is a popular, user-friendly standardized code for generating synthetic material instantiations. Subsequently some recent developments in modeling polyphase materials and polycrystalline materials, containing heterogeneities such as micro-twins and particulates, are discussed.

## 2 Creating Statistically Equivalent Virtual Polycrystalline Microstructures Using DREAM.3D

DREAM.3D is an open-source software package focused on creating a high-level programming environment to process, segment, quantify, represent, and manipulate digital microstructure data. A central goal of DREAM.3D is to enable the translation of microstructure quantification to a digital basis with easy-to-use software tools. The DREAM.3D environment is constructed to allow independently developed filters and plug-ins to interface with one another, enabling small research groups, government laboratories, start-up companies, and major industrial corporations to

collaborate and leverage each other's work. While DREAM.3D is a general suite of microstructure processing tools, one of the most common uses of the software is generating virtual material volumes for input into simulations of various types. This can readily be seen by viewing the references to the original publication introducing DREAM.3D (Groeber and Jackson 2014). The latest release of DREAM.3D can be downloaded from the website given in Jackson (2018).

Synthetic builders generally consist of two major processes, viz., generating features and spatial arrangements within a computational volume. Features may be generated by sampling the size, shape, and morphological and crystallographic orientation distributions observed by some experimental technique. Next, the features are placed in the volume with specific focus on the local neighborhoods of features. The sampling procedure, as well as the constraints used to place the features in the volume, is briefly described here. More detailed descriptions of the synthetic generation procedure can be found in Groeber et al. (2008a, b).

## 2.1 Representative Feature Generation

Idealized geometric objects have distributions of size, shape, and morphological orientation equivalent to those observed in the experimental volume, representing grains. In this representation, each feature is modeled as a geometric object (i.e., ellipsoid, super-ellipsoid, cube-octahedron, etc.) with a volume (V), a set of aspect ratios (b/a, c/a), and a morphological orientation ($\phi_1$, $\Phi$, $\phi_2$) corresponding to the orientation of the major principal axes (a,b,c) relative to the global axes. First the experimental feature volume distribution is sampled, which is represented by the cumulative distribution function (CDF) fit to the experimental data. Many investigations have shown the feature volume distribution to be well represented within $1 \to 2$ standard deviations of the mean, by a log-normal distribution (Zhang et al. 2004; Groeber et al. 2008b). It has been shown in Donegan et al. (2013) that grain size distributions tend to deviate from log-normal near the tails of the distribution. This can be accounted for by sampling from a piecewise distribution using a different form near the tails. Features are sampled until the total volume of all features generated is slightly larger than the volume of the synthetic microstructural model (typically around 10% larger). Additional volume is needed because some features may lie partially outside the domain or overlap other features. If the volume being generated is to have periodic boundaries, then additional volume is not needed because the portion of the features that fall outside of the domain will be placed on the opposite side. Subsequent to the volume assignment, feature shapes are assigned in conformity with CDFs of the aspect ratios (b/a, c/a). The corresponding CDFs are represented in terms of a beta distribution, due to its bounds of [0,1]. Additionally, the shape distributions are treated as a function of grain size by assigning unique shape distribution functions to discretely binned volume ranges of grain size. The morphological orientation of each feature is defined by a set of rotations ($\phi_1$, $\Phi$, $\phi_2$) needed to transform the global coordinates (X,Y,Z) onto the principal axes of the feature (X', Y', Z'). The orientation space is discretized into cubic bins, and the density in each bin represents the fraction of grains with that morphological

orientation. Ellipsoidal orientations are created and assigned based on sampling this probability density function (PDF), similar to the size and shape distributions. The output of this process is a set of geometric objects, representative of the features having statistically equivalent volume, aspect ratio, and morphological orientations as the experimental reference data.

## 2.2    Feature Placement

Spatial arrangement of features in a microstructure and their subsequent interactions drive local material response that can manifest in macroscopic heterogeneity. After generating a set of geometric objects, it is important that the placement of the features captures the local microstructural discontinuities. There are multiple issues to consider when packing the features. The density of the objects, represented by the features, is one of the largest factors in developing the packing algorithm. For example, features representing particles of a low volume fraction phase will certainly be placed differently than features representing grains in a fully dense polycrystalline material. In the fully dense grain, for example, care must be taken to pack the volume as densely as possible, but minimize overlap between features in order to retain each feature's prescribed shape. In both cases, the local neighborhood of the feature (i.e., neighboring features) must also be addressed during placement. The low volume fraction particles should be spaced equivalently to the experimental/reference data, and the densely packed grains should neighbor grains of sizes, shapes, and orientations similar to those seen in the experimental/reference data.

Alternative viable options for feature packing have been discussed in Groeber et al. (2008a) and Saylor et al. (2004). The approach used in DREAM.3D is a hybrid of methods in these references. The set of voxelized features are initially randomly placed in the volume. The features are then moved and swapped while enforcing constraints such as overlap or gap limits, number of neighboring features, size distributions of neighboring features etc. This approach generally yields near-optimal space filling through the overlap/gap limits and produces realistic neighborhoods by requiring local grain arrangements to match experimentally obtained metrics. The number of constraints affects the feasibility of finding a globally optimum arrangement. All clusters of unassigned voxels, corresponding to morphological incompatibility of features, are filled by a pseudo-grain coarsening process. A constrained Voronoi tessellation method discussed in Groeber et al. (2008b) is implemented to replace the voxelized representation of the grain aggregates by a solid-body surface representation of the grain boundaries.

A final step in the generation procedure is the assignment of crystallographic orientations to the placed features. The process of assigning crystallographic orientations is similar to the morphological orientation assignment process previously described, though they do not affect the grain morphology. Rotations transform the global coordinate axes to the crystal coordinate system, rather than the principal axes of the grain. Orientations are swapped and replaced while optimizing comparison to the experimental orientation and misorientation distributions.

## 2.3 Generating Statistically Equivalent 3D Virtual Microstructures

An example, demonstrating the use of the DREAM.3D software for microstructural characterization and 3D virtual microstructure reconstruction of a titanium alloy Ti-7Al, is shown in Fig. 1. This material has a predominantly *hcp* crystallographic structure (Pilchak 2013). A surface scan of the electron backscattered diffraction (EBSD) maps of the Ti-7Al alloy is shown in Fig. 1a. The morphological and crystallographic orientation, misorientation, and micro-texture distribution data from EBSD scans are input into the DREAM.3D software. The simulated statistically equivalent microstructure of dimensions $300 \times 300 \times 300\,\mu m$ containing 515 grains is shown in Fig. 1b with colors representing the $< c >$-axis misorientation with neighboring grains. The simulated pole figures are compared with those from EBSD data in Fig. 1c. Furthermore the probability density functions of misorientation and grain size are compared with those from EBSD maps in Fig. 1d, e respectively. Good agreement is generally seen between the simulated and experimental probability density functions.

## 3 Beyond DREAM.3D: Creating Statistically Equivalent RVEs of Polycrystalline and Polyphase Microstructures

While the DREAM.3D software is capable of generating virtual polycrystalline microstructures and SERVEs for various metals and alloys, incorporation of more complex microstructural features is still in nascent stages. Such features include those contained in polyphase and polycrystalline microstructures, e.g., annealing twins or particles and precipitates inside grains. This section will summarize a suite of algorithms that have been developed for creating M-SERVEs of complex polycrystalline and polyphase microstructures belonging to three distinct categories. These are:

1. Polycrystalline microstructures with localized features like annealing $\Sigma 3$ twin boundaries, shown in Fig. 2a;
2. Multiphase microstructures like subgrain $\gamma - \gamma'$ microstructure of Ni-based superalloys, shown in Fig. 7a;
3. Polycrystalline microstructures with dispersed precipitates and particles, e.g., for Al 7075-T6, shown in Fig. 10a.

## 3.1 Polycrystalline Microstructures with Annealing Twin Boundaries

Figure 2a shows EBSD images of a set of parallel sections of the superalloy René-88 DT microstructure, obtained by wire electrical discharge machining (EDM) of
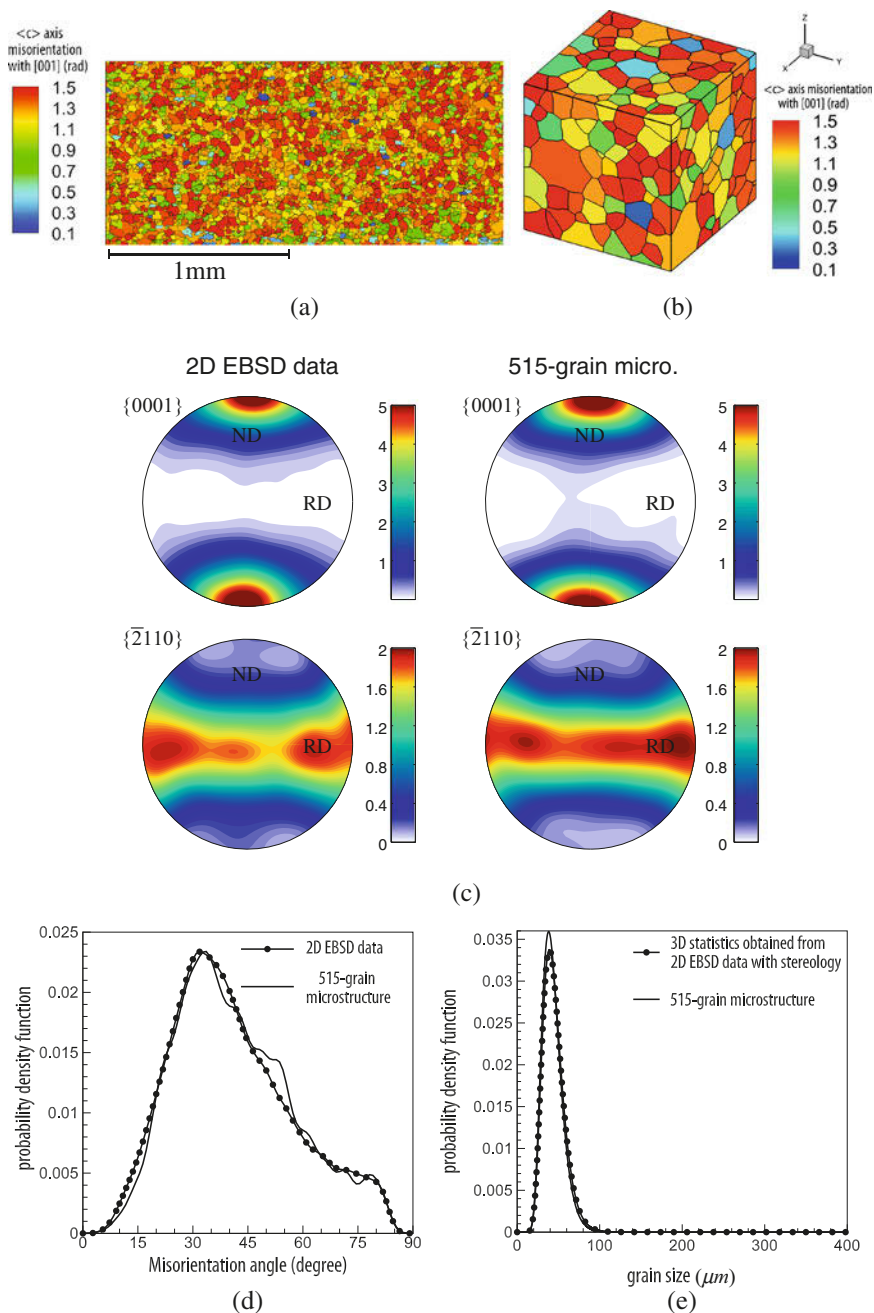
**Fig. 1** (**a**) EBSD scan of Ti-7AL; (**b**) statistically equivalent $300 \times 300 \times 300\,\mu m$ virtual microstructure containing 515 grains and showing the color plot of $< c >$-axis misorientation; comparison of (**c**) orientation distribution, (**d**) misorientation distribution, and (e) grain size distribution of the simulated microstructure with those from the EBSD data
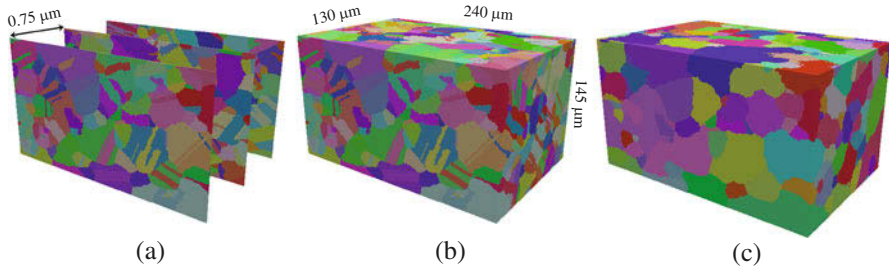
**Fig. 2** (**a**) Serial sectioned EBSD images, (**b**) computer-assembled sections of EBSD images manifesting polycrystalline microstructure including twins, and (**c**) polycrystalline microstructure of parent grains only after removing twins, for the Ni-based superalloy René88-DT

a $10 \times 5 \times 1$ mm sample (Lenthe 2017). The 3D assembly and microstructure reconstruction are performed in the DREAM.3D software using (i) slice registration and alignment, (ii) voxel level cleanup, (iii) feature segmentation, and (iv) artifact removal as detailed in Bagri et al. (2018). Twin-related domains are grouped with a $5°$ tolerance on both the disorientation axis and the disorientation angle. The resulting stacked and assembled 3D polycrystalline microstructure containing annealing twins is shown in Fig. 2b, where the grains are segmented with a $2°$ tolerance. The ensemble contains 440 twins in 300 parent grains for a total of 740 twins and grains. The polycrystalline microstructure is dominated by large aspect ratio, annealing $\Sigma 3$ twins that have a $60°$ misorientation angle about the $< 111 >$ crystal lattice axis. Details on the reconstruction of M-SERVEs are given in Bagri et al. (2018).

The following steps are executed in sequence to generate statistically equivalent volumes of twinned polycrystalline microstructures from scanned EBSD images.

1. Process the EBSD section data and construct the digitally assembled polycrystalline ensemble including twins;
2. Identify and remove twins from the digitally assembled microstructure to manifest the parent grains shown in Fig. 2c;
3. Extract the statistics of parent grains from the EBSD data;
4. Create statistically equivalent virtual parent grain microstructures from the 3D EBSD data;
5. Extract correlation statistics of twins with respect to parent grains from the EBSD data in the digitally assembled microstructure;
6. Insert twins in the parent microstructure to match statistical correlations.

The DREAM.3D software (Groeber and Jackson 2014) is employed in steps 1–4 to create the digital polycrystalline ensembles from EBSD data in Fig. 2d: remove twins in Fig. 2c and subsequently extract statistics of the parent grains. The statistics of characteristic features in the twin-free parent grains in Fig. 2c, including probability distributions of grain size, orientation, and misorientation
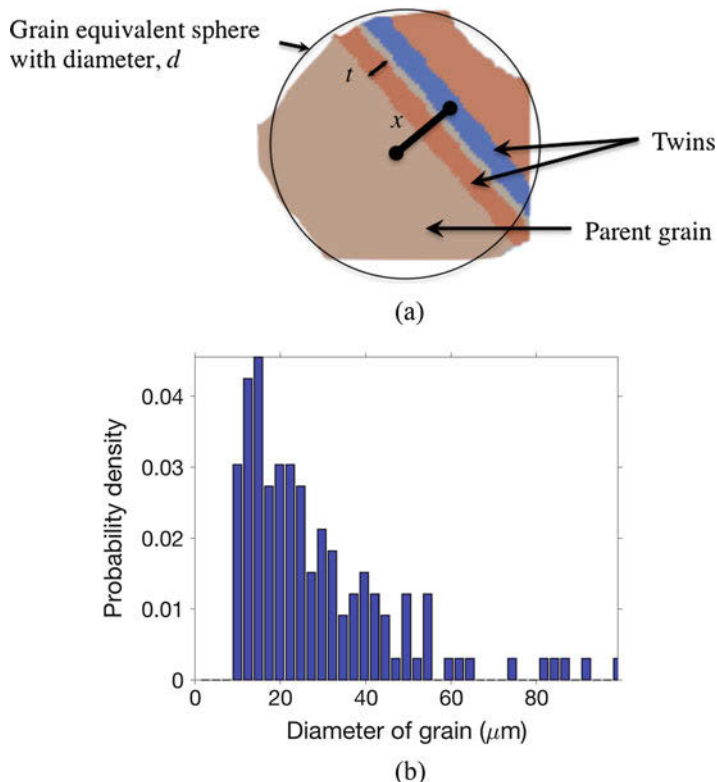
**Fig. 3** (**a**) Schematic of twins and parent grain and (**b**) statistics of grain size distribution after removing twins

angle, are acquired following (Groeber et al. 2008a, b). The probability density distribution of grain size is shown in Fig. 3b. For instances where only 2D EBSD are available, methods of estimating 3D grain sizes from 2D surface data, e.g., through the stereology relation $d_{3D} \sim 4/\pi d_{2D}$ (Groeber et al. 2008a), may be used.

In step 5, sample statistics are extracted from the EBSD data to generate probability distribution and correlation functions of twins with respect to parent grains. These are subsequently used to insert twins in the parent microstructure. A schematic representing the relation of a twin with the parent grain is shown in Fig. 3a. Statistical analysis shows the parent grain size $d$, number of twins in parent grain $n$, minimum distance $x$ of the twin from the parent grain centroid, and the twin thickness $t$ are strongly correlated. The joint probability distribution of the correlation between twins and parent grains is expressed through a function $P_0(d, n, t, x)$. The correlation statistics and joint probability distributions are shown in Fig. 4.

The algorithm to insert twins in the parent grain microstructure consists of the following steps.
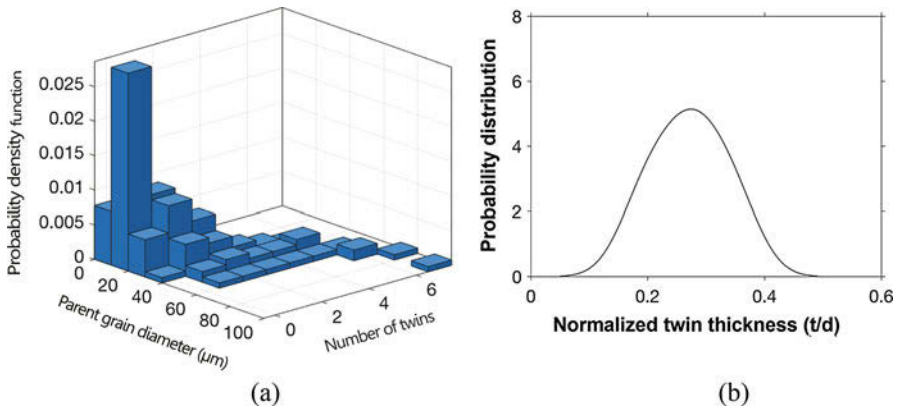
**Fig. 4** Correlation statistics from the EBSD data: (**a**) joint probability density distribution of the parent grain size and number of twins and (**b**) conditional probability density distribution of twin thickness

1. Use the joint probability distribution of parent grain size and number of twins $P_1(d, n) = \iint P_0(d, n, t, x) \, dt \, dx$ and the conditional probability distributions of both twin distance from parent centroid $P_2(x | d = D, n = N)$ and twin thickness $P_3(t | d = D, n = N)$ for twin insertion.
2. Determine the number of twins, twin thickness, and the twin distance from parent centroid using a Monte Carlo-based acceptance-rejection scheme.
3. Locate the (111) plane at a distance $x$ from the parent centroid.
4. Identify voxels that are at a distance $\frac{t}{2}$ from the mid-thickness (111) plane.
5. Calculate the rotation matrix from the rotation matrices of the parent grain and the twin with respect to parent, i.e., $\mathbf{R} = \mathbf{R}_{\text{parent}} \mathbf{R}_{\text{twin}}$
6. Determine Euler angles of the twin, and reassign them to voxels of the twin.

The four-dimensional probability distribution $P_0(d, n, t, x)$ requires a very large number of grains. Hence, a marginal probability function is used, when a smaller set of statistical information is available. In step 1, the joint probability density distribution, e.g., in Fig. 4a, is used for parent grain and number of twins per parent, while the conditional probability distributions are used for the twin thickness (Fig. 4b) and twin distance from the parent centroid. With this assumption, the four-dimensional distribution space is approximated as:

$$P_0(d, n, t, x) \approx P_1(d, n) P_2(x | d = D, n = N) P_3(t | d = D, n = N) \qquad (1)$$

This approximation is valid for any parent grain size and associated twins, where the twin size and distance from the parent centroid are uncorrelated. When inserting the coherent twins, they must be placed with the proper orientation relationship to the parent. The twin boundary plane orientation in the specimen frame $\mathbf{m}_s$ is first determined using the crystallographic orientation vector of the parent grain as
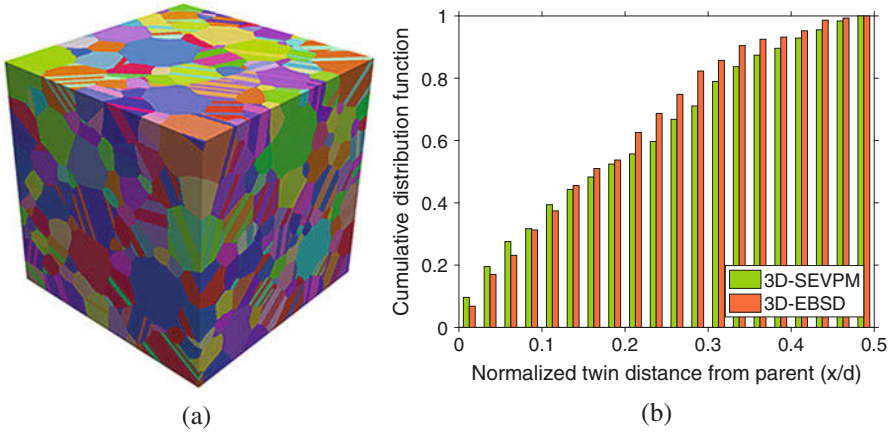
**Fig. 5** (**a**) Simulated 3D statistically equivalent polycrystalline microstructure (3D-SEVPM) with twins inserted in the parent grains and (**b**) comparison of the cumulative distribution function of twin distance $d$ from parent centroid in the EBSD scan and 3D-SEVPM of size $250\,\mu$m

$\mathbf{m}_s = \mathbf{R}_{\text{parent}}\mathbf{m}_p$, where $\mathbf{R}_{\text{parent}}$ and $\mathbf{m}_p = (111)$ are the parent grain rotation matrix and grain boundary plane orientation vector in the crystal coordinates, respectively. Subsequently, the plane is located at a previously selected distance $x$ from the parent centroid. The grain boundary plane will pass through the point $\mathbf{x}_p = \mathbf{x}_c + x\frac{\mathbf{m}_s}{\|\mathbf{m}_s\|}$. Here, $\mathbf{x}_p$ is the position vector of a point in the grain boundary, $\mathbf{x}_c$ is the position vector of the parent grain centroid, and $\|\mathbf{m}_s\|$ is the norm of the grain boundary normal. The rotation matrix of the twin in the specimen frame $\mathbf{R}$ is obtained from the rotation matrix of parent grain $\mathbf{R}_{\text{parent}}$ and the rotation matrix of the twin with respect to parent grain $\mathbf{R}_{\text{twin}}$ as $\mathbf{R} = \mathbf{R}_{\text{Parent}}\mathbf{R}_{\text{twin}}$. Using the components of $\mathbf{R}$, the Euler angles of twins are expressed as:

$$\phi_1 = tan^{-1}(-\frac{R_{13}}{R_{23}}), \quad \Phi = cos^{-1}(R_{33}), \quad \phi_2 = tan^{-1}(\frac{R_{31}}{R_{32}}) \tag{2}$$

Repeating steps 1–6, the set of twins are inserted into the parent grain microstructure.

### 3.1.1 An Example of Validated M-SERVE Generation

The algorithm is used with EBSD data in Alam et al. (2016) for validating the virtual microstructure generation process. The statistics shown, e.g., in Fig. 4 are used to insert twins in the parent grain microstructure. A $250 \times 250 \times 250\,\mu$m 3D-SEVPM consisting a total of 1700 parent and twins is shown in Fig. 5a. The cumulative distribution function of the twin distance $d$ from parent centroid for this 3D-SEVPM is compared with that from the EBSD scan in Fig. 5b.

Studies in Bagri et al. (2018) have shown that the 3D-SEVPM converges to the M-SERVE at $150\,\mu$m, which corresponds to approximately 400 grains and
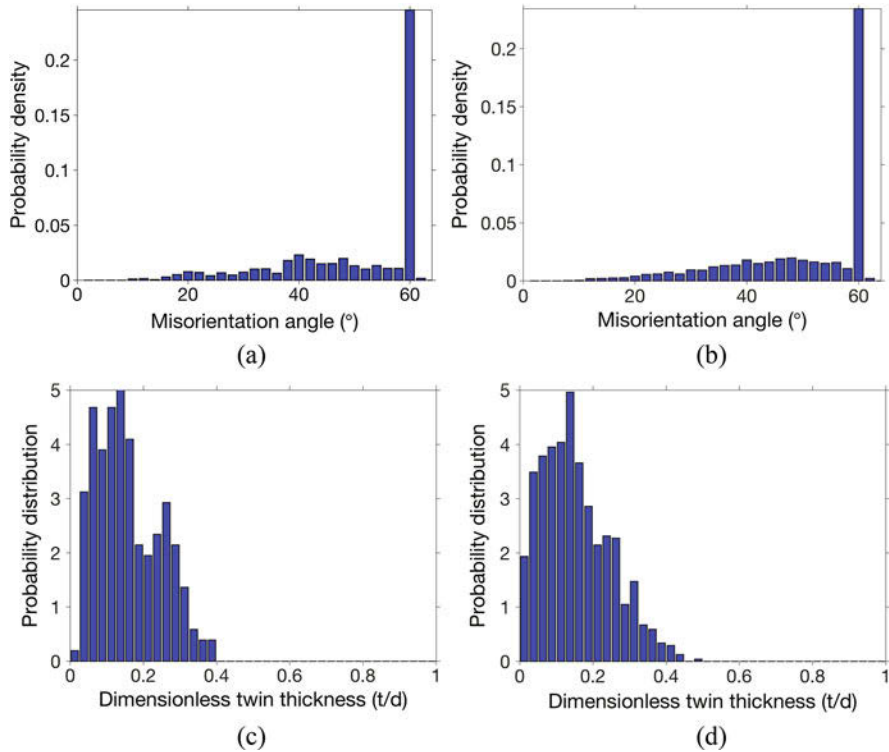
**Fig. 6** Probability distribution of (**a**, **b**) misorientation angle and (**c**, **d**) twin thickness from the EBSD data and the M-SERVE, respectively

twins. Comparison of the M-SERVE statistics with the EBSD data is made through a few probability distributions. The distribution of the global misorientation and twin thickness is compared in Fig. 6. The prominent peak in misorientation angle distribution at 60° indicates the presence of twins in the microstructure. A good agreement is obtained for all the distribution plots. This is a step in validating the virtual microstructure generation method. In both the EBSD data and M-SERVE, about 40% of the parent grains are seen to remain untwinned.

## 3.2    Two-Phase Microstructures Underlying Polycrystalline Grains

Polycrystalline nickel-based superalloys like René-88 DT have an underlying subgrain-scale two-phase microstructure consisting of a dispersion of $\gamma$ and $\gamma'$ precipitates, as shown in Fig. 7a. Subgrain-scale morphological characteristics like volume fraction, mean size, and channel-width or spacing of $\gamma'$ precipitates have a major effect on their mechanical properties (Unocic et al. 2011). Modeling
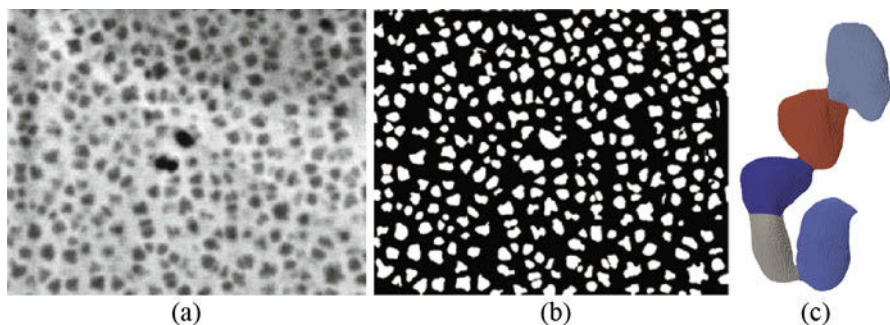
**Fig. 7** (**a**) Scanning electron microscope image of a microstructural section of Ren*é*-88 DT acquired by FIB serial sectioning, (**b**) segmented SEM image after thresholding and despeckling, and (**c**) spurious connectivity from serial sectioning after segmentation

their mechanical and physical behavior requires robust representation of these morphological features in the M-SERVE. Many approaches have been employed to include precipitate structure in performance simulations, e.g., Pollock and Argon (1992), Nouailhas and Cailletaud (1996), Busso et al. (2000), Fromm et al. (2012), Parthasarathy et al. (2004) and Keshavarz and Ghosh (2015).

This section discusses a methodology developed in Pinz et al. (2018) for generating M-SERVEs of two-phase $\gamma-\gamma'$ microstructures from 2D microstructural scans of 3D data, using the following steps.

1. *FIB-SEM serial sectioning*: Acquisition of high-fidelity 3D $\gamma-\gamma'$ microstructural data necessitates a high-throughput automated serial-sectioning process coupled with high-resolution SEM data extraction.
2. *Image processing and data cleanup*: Cleanup and subsequent segmentation of $\gamma$ and $\gamma'$ phases generate a reference 3D voxelization of the microstructure.
3. *Feature extraction and statistical characterization*: Parametrization and statistical characterization of the $\gamma'$ precipitate morphology and designation of the relative precipitate positions are needed for establishing spatial distributions.
4. *Statistically equivalent microstructural reconstruction*: Optimally minimum microstructures are generated with morphological and spatial statistics equivalent to those of the large-imaged microstructures.

### 3.2.1 Data Extraction with 3D FIB Serial Sectioning and SEM-Based Imaging

Recent advances in tomographic methods, e.g., in Uchic et al. (2006), and Echlin et al. (2014) and image processing tools have greatly increased the accessibility of 3D data sets for a variety of materials. Microstructural data used for the generation of the polyphase M-SERVE is obtained from Lenthe (2017), where a FIB is used to expose parallel layers of the material that are imaged with a SEM. A sample section is shown in Fig. 7a. The in-plane resolution is 2.5 nm per pixel with 20 nm between

the slices. The sections are aligned via a convolution method to reduce the effects of instrumental drift. The contrast difference between the $\gamma$ and $\gamma'$ phases is not large enough to allow for robust automatic segmentation of the SEM images (Pollock and Tin 2006) and can require extensive cleanup.

### 3.2.2   Image Segmentation

The FIB-SEM technique in Lenthe (2017) is used to collect 180 gray-scale images, each with a size of 1996 × 1596 pixels. The resultant image stack yields a 3 × 4 × 5 µm volume. An automated image segmentation process begins with local noise reduction followed by image sharpening to enhance the contrast between the phases. The sharpened image stack is segmented with a minimum entropy threshold (Li and Lee 1993) as shown in Fig. 7b. Sections are interpolated between extracted images in all three directions. The interaction volume of the SEM causes spurious connectivity between the precipitates as shown in Fig. 7c. A watershed segmentation procedure (Meyer 1994) is used to separate the conjoined precipitates using a gradient field given as:

$$
\mathcal{G}(i, j, k) = \mathcal{B}_{\text{int}}(i, j, k) \frac{\sum_{\bar{i}=0}^{N_x} \sum_{\bar{j}=0}^{N_y} \sum_{\bar{k}=0}^{N_z} \mathcal{D}_{\text{int}}(\bar{i}, \bar{j}, \bar{k}) e^{-\left[\left(\frac{(\bar{i}-i)^2}{2\sigma^2}\right)+\left(\frac{(\bar{j}-j)^2}{2\sigma^2}\right)+\left(\frac{(\bar{k}-k)^2}{2\sigma^2}\right)\right]}}{\sum_{\bar{i}=0}^{N_x} \sum_{\bar{j}=0}^{N_y} \sum_{\bar{k}=0}^{N_z} e^{-\left[\left(\frac{(\bar{i}-i)^2}{2\sigma^2}\right)+\left(\frac{(\bar{j}-j)^2}{2\sigma^2}\right)+\left(\frac{(\bar{k}-k)^2}{2\sigma^2}\right)\right]}}
$$

$$(3)$$

where $\bar{i}, \bar{j}, \bar{k}$ are dummy indices, $N_x$, $N_y$, and $N_z$ are the number of voxels in the $x, y, z$ directions, respectively, and $N_{\text{slice}}$ is the total number of slices after interpolation. $\mathcal{D}_{\text{int}}$ is a map from each voxel to the value of its distance to the nearest boundary in the plane, and $\mathcal{B}_{\text{int}}$ is a binarized map relative to $\mathcal{D}_{\text{int}}$. The latter is an indicator function of whether a voxel $(i, j, k)$ is in a precipitate or not. The standard deviation $\sigma$ of the Gaussian blur is set to $\frac{1}{10}$th of the mean particle radius. This gradient field is chosen to reduce disconnected over-segmentation by the watershed algorithm. The purpose of the 3D Gaussian blur is to mitigate effects from the voxelization of the precipitate edges. After application to the initially connected inclusions, the watershed segmentation algorithm produces a final binary voxelized map $\mathcal{B}_{\text{final}}$ that contains approximately 6000 contiguous precipitates.

### 3.2.3   Mapping Precipitate Domain to a 3D Parametric Function Model

Homogenized constitutive models require the precipitate morphology to be described by a parametrized function with a finite number of parameters and coefficients. These parameters may be calibrated from the actual surface profiles by optimization. The chosen parametric function for surface representation represents a generalized super-ellipsoid (GSE), delineated as:

$$
\left(\frac{\bar{x}}{a}\right)^{N1} + \left(\frac{\bar{y}}{b}\right)^{N2} + \left(\frac{\bar{z}}{c}\right)^{N3} = 1
$$

$$(4)$$

The position vector $\bar{\mathbf{x}}(= \bar{x}, \bar{y}, \bar{z})$ corresponds to the location of a GSE surface point $\mathbf{x}_p(= x_p, y_p, z_p)$ relative to its centroid $\mathbf{x}_0(= x_0, y_0, z_0)$ in its principal coordinate system represented by the Euler angles $(\phi_1, \Phi, \phi_2)$. The relative coordinates are expressed as $\{\bar{\mathbf{x}}\} = [\mathbf{R}]\{\mathbf{x}_p - \mathbf{x}_0\}$, where $[\mathbf{R}]$ is the rotation matrix for the precipitate coordinate system. In Eq. (4) $a, b, c$ correspond to the principal axis lengths of the super-ellipsoid, and $N1, N2, N3$ are exponents representing the shape. The parametrized function in Eq. (4) entails evaluation of the parameter set $Y_{\text{par}} \in \{x_0, y_0, z_0, N1, N2, N3, a, b, c, \phi_1, \Phi, \phi_2\}$. The parameters are calibrated by solving an optimization problem that minimizes the orthogonal distance between $m$ surface points of a precipitate and the parametrized surface as:

$$\underset{Y_{\text{par}}}{\text{Minimize}} \sum_{i=1}^{m} D_i^2 = \sum_{i=1}^{m} \|\mathbf{x}_i - (\mathbf{x}_p)_i\|^2 = \sum_{i=1}^{m} \left\{ x_i - (x_p)_i \right\}^2$$

$$+ \left\{ y_i - (y_p)_i \right\}^2 + \left\{ z_i - (z_p)_i \right\}^2 \tag{5a}$$

subject to the constraint that each point $i$ belongs to the GSE surface

$$\left( \frac{\bar{x}_i}{a} \right)^{N1} + \left( \frac{\bar{y}_i}{b} \right)^{N2} + \left( \frac{\bar{z}_i}{c} \right)^{N3} = 1 \ \ \forall \ i \in [1, m] \tag{5b}$$

The variable $D_i$ corresponds to the absolute Euclidean distance between a surface point on the precipitate at coordinates $(\mathbf{x}_i = (x_i, y_i, z_i))$ for a point $i$ and its conjugate surface point $(\mathbf{x}_p)_i = (x_p)_i, (y_p)_i, (z_p)_i$ on the parametrized GSE in the current iterate.

### 3.2.4 Validation of Precipitate Mapping and Reconstruction

Distributions of morphological parameters of the precipitates generated are used to validate the effectiveness of the parametrization.

*I. Dice Index:* The Dice index (DI) is used as a goodness-of-fit metric to compare the orthogonal distance minimization (ODM) algorithm with an alternate moment-based algorithm proposed in MacSleyne et al. (2009). DI is defined as the volume of overlap between two objects (the actual precipitate and the GSE) divided by the composite volume of the union of the two objects as: $DI = \frac{V_{FIB} \cap V_{GSE}}{V_{FIB} \cup V_{GSE}}$. It ranges from 0 corresponding to no overlap to 1 for perfect intersection. In Fig. 8a, the distribution of DI is compared for GSE's generated by alternate methods. "FULL" corresponds to $N1 \neq N2 \neq N3$, while "REDUCED" has $N = N1 = N2 = N3$.

*II. Size, Shape, and Orientation Distributions:* The distributions of the major, minor, and intermediate axes $a$, $c$, and $b$, respectively, represent the size and aspect ratio of the precipitate. Figure 8b compares the distribution of the shape parameter $N$ obtained in the ODM algorithm to a log-normal distribution representation of the same by the maximum likelihood estimation (MLE). Orientations of the GSEs by the ODM algorithm with reduced shape parameters are used to generate equivalent orientation distribution functions using spherical harmonics for crystallographic texture.
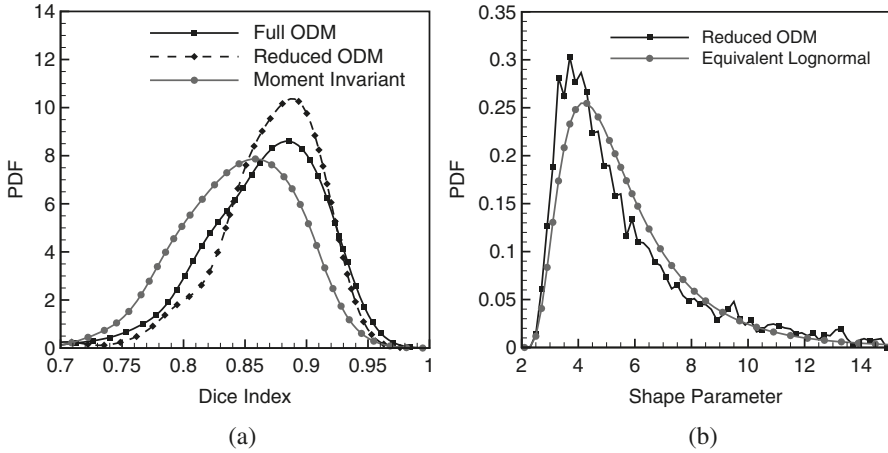
**Fig. 8** Probability distribution of (**a**) the DI for equivalent ellipsoids generated by alternative methods and (**b**) shape parameters in the ODM algorithm and MLE shifted log-normal distribution

### 3.2.5   Statistically Equivalent Microstructure by Dispersing Precipitates

A set of precipitates with representative morphological parameters are generated to yield a desired volume fraction. A placement algorithm that involves random allocation with local perturbation to avoid precipitate overlap, as well as microstructural shuffling through an energy minimization scheme for positional stability, is implemented. This method is iteratively continued with precipitate size scaling until the experimental volume fraction is attained.

An optimization schedule is executed to minimize the difference in the two-point correlation function $S_2$ for the experimental microstructure and the 3D statistically equivalent virtual microstructures (3D-SEVMs). The two-point correlation function $S_2$ is a statistically convergent measure of the microstructural heterogeneity in Tewari et al. (2004) and Jiao et al. (2007). For isotropic distributions, it can be approximated by a closed-form solution as:

$$S_2(r) = V_f{}^2 + V_f(1 - V_f)e^{\frac{-r}{r_0}} \frac{\sin\left(\frac{2\pi r}{a_0}\right)}{\frac{2\pi r}{a_0}} \tag{6}$$

where $V_f$ represents the volume fraction of precipitates and $a_0$ and $r_0$ are calibrated parameters. A genetic algorithms (GA)-based optimization (Goldberg 1989) is used to minimize the difference in the $S_2$ function. The fitness function is expressed as:

$$FF_{S_2} = \left(\frac{a_o - a_o^{\text{target}}}{a_o^{\text{target}}}\right)^2 + \left(\frac{r_o - r_o^{\text{target}}}{r_o^{\text{target}}}\right)^2 \tag{7}$$
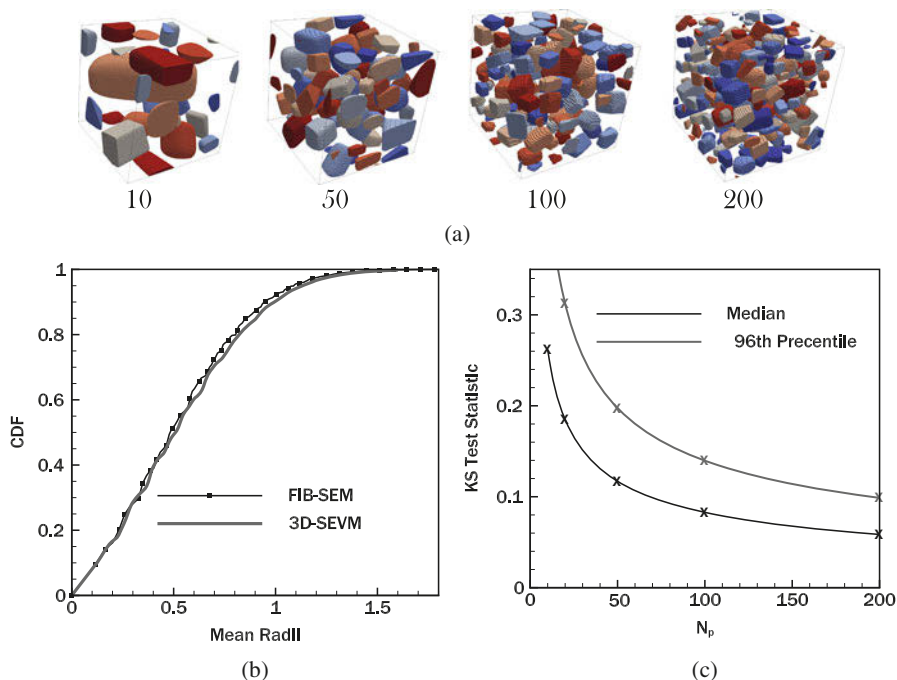
**Fig. 9** (**a**) Reconstructed 3D-SEVMs with 10, 50, 100, and 200 precipitates; (**b**) cumulative distribution function of distance to precipitate surface, generated from experimental data and 3D-SEVM; and (**c**) median and 96th percentile expected KS test statistic with 10, 50, 100, and 200 precipitates

Figure 9a shows examples of the reconstructed 3D-SEVMs for $N_p = 10, 50, 100,$ and 200 precipitates. Validation tests of 3D-SEVMs are conducted by comparing morphological metrics that are not optimized during the reconstruction process. One relevant metric for plastic deformation is the *distance to precipitate surface* (DPS) distribution. Figure 9b shows a cumulative distribution function of DPS for experimental and 3D-SEVM volumes, exhibiting good agreement. Figure 9c plots the Kolmogorov-Smirnov (KS) test (Massey 1951) statistic between the 3D-SEVM with $N_p = 10, 50, 100,$ and 200 precipitates and the FIB-SEM microstructural data. Both the median and an upper bound of the sampling error with frequency of 0.96 are plotted in this figure. The 3D-SEVM with 200 precipitates is sufficient for convergence, and hence this is the designated M-SERVE as detailed in Pinz et al. (2018).

## 3.3   Polycrystalline Microstructures with Dispersed Precipitates

Many engineering alloys have precipitates or particles dispersed in their polycrystalline microstructure. For example, the 7000-series aluminum alloys,
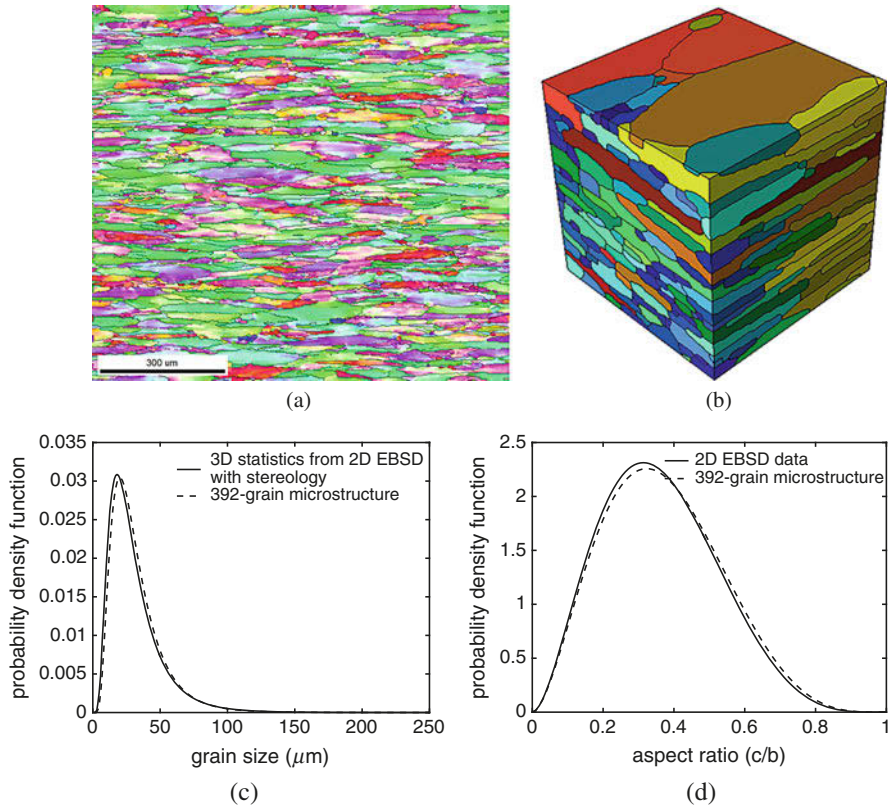
**Fig. 10** Inverse pole figure maps of EBSD data and SEM images of TD/ND plane of the Al7075-T651 sample containing particles in polycrystalline microstructure; (**b**) DREAM.3D reconstructed 392-grain virtual microstructure; comparing statistical distributions of the virtual grain morphology with EBSD data for (**c**) grain size, (**d**) aspect ratio c/b

e.g., Al 7075-T6, contain iron-rich or magnesium-rich precipitates in the aluminum matrix, as shown in Fig. 10a. Various models have been proposed for microstructure generation of porous materials and particle-reinforced metals, e.g., in Baniassadi et al. (2011), Rollett et al. (2006) and Guo et al. (2014). Following developments in the previous sections, this section discusses a method for constructing 3D virtual microstructures from 2D micrographs, accounting for grain and particle shape distributions, spatial arrangements of precipitates, as well as precipitate-grain spatial correlations. Stereological concepts are numerically implemented into a genetic algorithm (GA)-based optimization framework as detailed in Tu et al. (2019).

### 3.3.1   Stereological Reconstruction of the Grain Microstructure

Figure 10a shows a representative EBSD scan of a cross section from an aluminum alloy (7075-T6). Crystallographic orientation and misorientation distributions, as
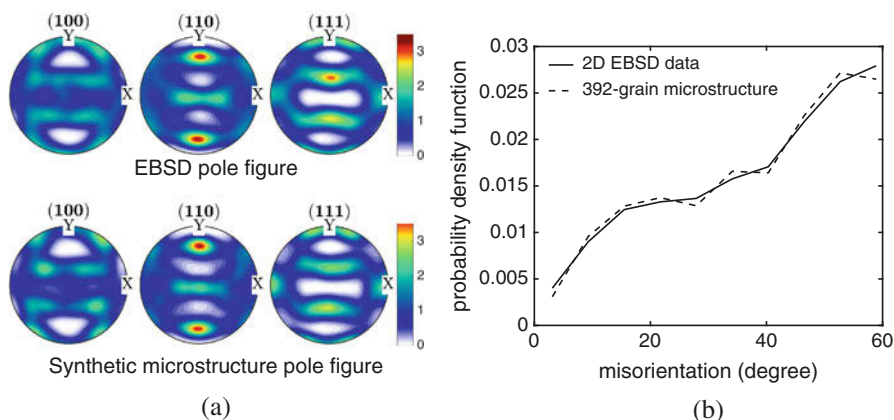
**Fig. 11** Comparing crystallographic distributions of the virtual grain with EBSD data for (**a**) orientation distribution (pole figures) and (**b**) misorientation distribution

well as morphological distributions, e.g., grain size, aspect ratio distributions, are extracted from these scans. Stereological estimations of 3D size and aspect ratio distributions are made from 2D surface data following (Underwood 1972), and the 3D microstructure is constructed by DREAM.3D using these distributions. Figures 10 and 11 compare the morphological and crystallographic statistics of a 392-grain virtual microstructure with those from EBSD data.

### 3.3.2    Mapping Distributions of Precipitates to Parametrized GSEs

SEM images of precipitates are mapped to distributions of 3D generalized super-ellipsoids (GSEs), similar to Sect. 3.3. Image processing of the SEM images includes contrast enhancement (adaptive histogram equalization), binarization (gray level thresholding), and noise removal (Wiener filtering). The SEM particle contours on the 2D surface are fitted to generalized super-ellipse using the 2D version of Eq. (5). The microstructures show precipitate clustering along the rolling direction (RD) (Rollett et al. 2007). The isotropic two-point correlation function or radial distribution function in Jiao et al. (2007) and Wang et al. (2016) is unable to characterize this directional clustering. The two-point correlation analysis of precipitates in the RD/ND plane in Fig. 12a shows a clear trend of directional clustering along the RD direction. Furthermore, the precipitate-grain spatial correlation is represented by the conditional probability distribution of the minimum distance to grain boundaries for a given precipitate size. Fracture toughness of aluminum alloys is sensitive to the closeness of precipitates to grain boundaries (Cai et al. 2007), which is obtained by overlaying the EBSD and SEM 2D data sets. Figure 12b manifests the clustering of precipitates of various sizes near grain boundaries.

### 3.3.3    GA-Based Stereological Mapping from 2D to 3D Microstructures

A genetic algorithm (GA)-based optimization method (Goldberg 1989) is used to search for parameters of the 3D precipitate distributions, by minimizing the
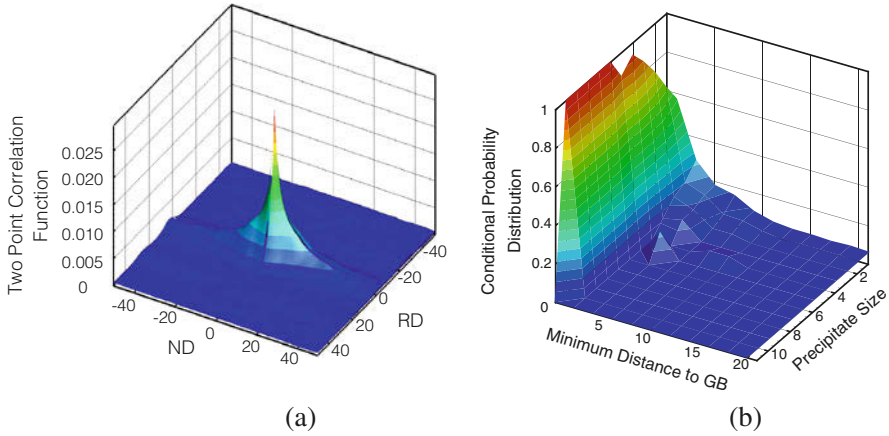
**Fig. 12** (**a**) Anisotropic two-point correlation function for precipitates and (**b**) conditional probability distribution of the distance to grain boundaries and precipitate size in the RD/ND plane

difference in cumulative distribution functions (CDFs) of 2D statistics of the virtual and experimental microstructures, using the equation:

$$\underset{X_k}{\text{Minimize}} \frac{1}{N_{\text{descriptor}}} \sum_{k=1}^{N_{\text{descriptor}}} \frac{1}{N_{\text{bin}}^{(k)}} \sum_{i=1}^{N_{\text{bin}}^{(k)}} \left| CDF_i^{\text{experimental}} - CDF_i^{\text{virtual}} \right| \tag{8}$$

where $N_{\text{descriptor}}$ corresponds to the number of descriptors, e.g., precipitate size, aspect ratio, shape parameter, and principal axes orientation. $X_k$ stores the 3D shape distribution parameters for the $k$-th descriptor, and $N_{\text{bin}}^{(k)}$ is the number of bins allocated to represent the cumulative distribution of descriptor $k$. Candidate sets of 3D super-ellipsoid semiaxis length $(a, b, c)$, shape parameter $(n)$, and principal axes orientation $(\alpha, \beta, \gamma)$ distribution parameters are stored in the array **X**. For every candidate set, representative 2D orthogonal sections are derived statistically from 3D virtual microstructures and compared with the EBSD surface image data. Subsequently, the parameters of the sectioned super-ellipses are solved with the method of undetermined coefficients. With known statistics of the virtually sectioned super-ellipses and experimental surface data, individual fitnesses are calculated from the minimization problem involving crossover and mutation operations.

### 3.3.4   Planting Precipitates in the DREAM.3D-Generated Grain Microstructure

The 3D super-ellipsoidal precipitates are now spatially dispersed in the 3D parent polycrystalline matrix. The dispersion of precipitate centroids is optimized by the GA methodology such that the anisotropic $S_2$ function and precipitate-grain spatial correlations of the sections match the experimental distributions. The fitness function of the optimization process is written as:

$$\text{Minimize} \atop \mathbf{C} \quad \frac{1}{N_{\text{descriptor}}} \sum_{k=1}^{N_{\text{descriptor}}}$$

$$\left\{ \frac{1}{N_{\text{section}}} \sum_{m=1}^{N_{\text{section}}} \frac{1}{N_{\text{bin}}^{(k)} * M_{\text{bin}}^{(k)}} \sum_{j}^{N_{\text{bin}}^{(k)}} \sum_{i}^{M_{\text{bin}}^{(k)}} \left| CDF_{ij}^{\text{virtual}} - CDF_{ij}^{\text{experimental}} \right| \right\}$$

$$(9)$$

where $\mathbf{C}$ is the centroidal coordinates of precipitates, $N_{\text{descriptor}}$ is the number of descriptors, $N_{\text{section}}$ is the number of sections, and $N_{\text{bin}}^{(k)}$ and $M_{\text{bin}}^{(k)}$ are the number of bins used for the cumulative distribution function of a descriptor $k$.

The resulting two-phase reconstructed microstructure is shown in Fig. 13b. Figure 13b shows that the precipitates cluster along the RD (x-axis) in good agreement with the experimentally obtained microstructural distribution. Quantitative comparison of virtual and experimental microstructures in Fig. 14 shows good match of the 2D precipitate-precipitate and precipitate-grain boundary correlations.

### 3.3.5 Convergence of the M-SERVE

The convergence of various microstructural descriptors is studied for assessing the M-SERVE size. The descriptors include grain morphology, grain crystallography, particle morphology, and particle-grain spatial correlation. The errors in these distributions are estimated by the Kolmogorov-Smirnov test. From the M-SERVE convergence characteristics in Fig. 15 for two-point and precipitate-grain correlations, it is found that the M-SERVE size is controlled by precipitate-grain correlations. It converges for approximately 450 grains with 4500 precipitates.



(a) (b)
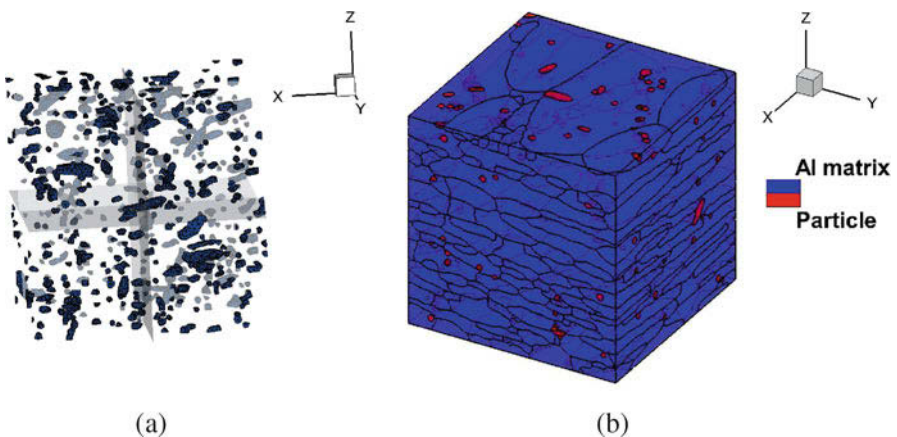
**Fig. 13** Representation of the of reconstructed two-phase microstructure with (**a**) precipitates only and (**b**) precipitates embedded in Al polycrystalline microstructure with grain boundaries
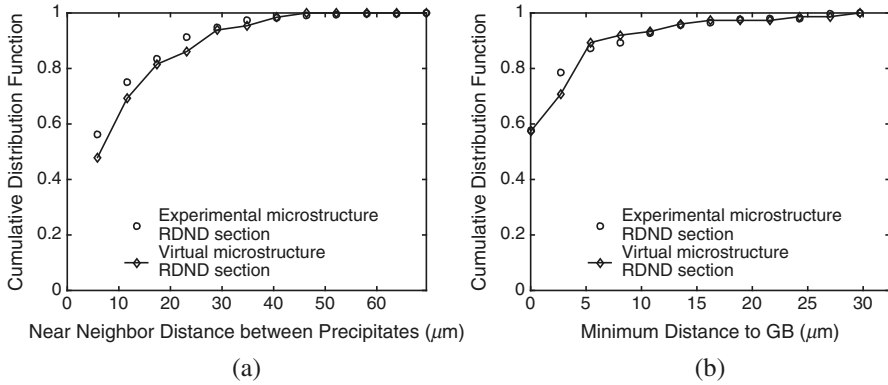
**Fig. 14** Convergence of 2D statistics of particle descriptors to the experimental data: (**a**) particle-particle near neighbor distance with respect to the RDND plane data, (**b**) particle minimum distance to GB in (RDND)
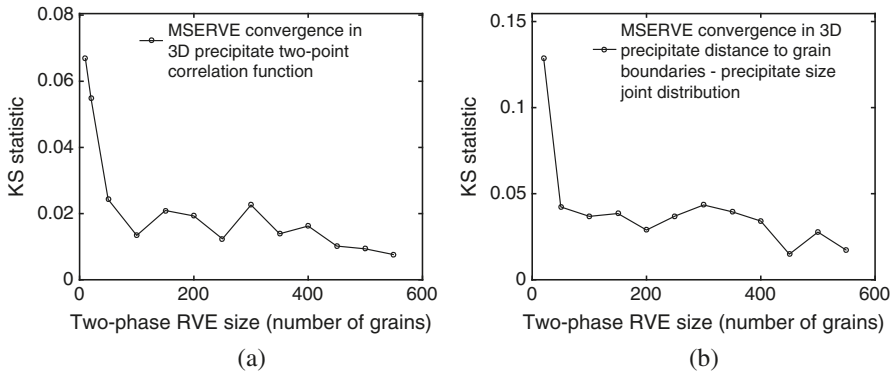


**Fig. 15** Convergence of M-SERVE with respect to (**a**) 3D particle two-point correlation function and (**b**) 3D distance to grain boundary-precipitate size joint distribution

## 4 Conclusions

This chapter discusses the development of computational methods for simulating statistically equivalent virtual microstructures of materials with complex hetero-geneities. The methods consider polycrystalline materials containing localized features such as annealing twins, particulates or precipitates, and subgrain-scale precipitates in their polycrystalline structure. Data from image analysis and charac-terization are used to construct 3D statistical distribution and correlation functions. This serves as input to the virtual microstructure generation process. 3D experimen-tal data extraction techniques are sometimes unavailable or prohibitively expensive to generate. Consequently, the methods accommodate computational approaches that infer 3D statistical descriptors and functions from 2D observations and surface

data, from stereology and other optimization-based projection techniques. The chapter introduces the open-source software package DREAM.3D that is now widely used for generating virtual microstructures of polycrystalline materials. It then goes beyond DREAM.3D into more newly developed methods for incorporating twins, particles, and subgrain-scale phases in polycrystalline microstructures. Finally the concept of the microstructure-based SERVE or *M-SERVE*, in which morphological and crystallographic characteristics of the microstructure are determinants of the statistically equivalent representative volume element, is introduced.

# References

Alam A, Eastman D, Jo M, Hemker KJ (2016) Development of a high-temperature tensile tester for micromechanical characterization of materials supporting meso-scale ICME models. JOM 11(68):2754–2760

Bagri A, Weber G, Stinville JC, Lenthe W, Pollock T, Woodward C, Ghosh S (2018) Microstructure and property-based statistically equivalent representative volume elements for polycrystalline Ni-based superalloys containing annealing twins. Met Mat Trans A49:5727–5744

Baniassadi M, Garmestani H, Li DS, Ahzi S, Khaleel M, Sun X (2011) 3-phase solid oxide fuel cell anode microstructure realization using 2-point correlation functions. Acta Mat 59(1):30–43

Bhandari Y, Sarkar S, Groeber MA, Uchic MD, Dimiduk D, Ghosh S (2007) 3D polycrystalline microstructure reconstruction from FIB generated serial sections for FE Analysis. Comput Mat Sci 41:222–235

Busso E, Meissonier F, O'Dowd N (2000) Gradient-dependent deformation of two-phase single crystals. J Mech Phys Sol 48:2333–2361

Cai B, Adams B, Nelson T (2007) Relation between precipitate-free zone width and grain boundary type in 7075-T7 Al alloy. Acta Mat 55(5):1543–1553

Donegan SP, Tucker JC, Rollett A, Barmak K, Groeber MA (2013) Extreme value analysis of tail departure from log-normality in experimental and simulated grain size distributions. Acta Mat 61(15):5595–5604

Echlin MP, Lenthe WC, Pollock TM (2014) Three-dimensional sampling of material structure for property modeling and design. Integ Mat Manuf Innov 3(1):21

Fromm BS, Chang K, McDowell DL, Chen L, Garmestani H (2012) Linking phase-field and finite-element modeling for process-structure-property relations of a Ni-base superalloy. Acta Mat 60:5984–5999

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading

Groeber MA, Jackson M (2014) DREAM.3D: a digital representation environment for the analysis of microstructure in 3D. Integr Mat Manuf Innov 3:5

Groeber MA, Haley BK, Uchic MD, Dimiduk D, Ghosh S (2006) 3D reconstruction and characterization of polycrystalline microstructures using a FIB-SEM system. Mat Char 57(4–5):259–273

Groeber MA, Ghosh S, Uchic MD, Dimiduk D (2008a) A framework for automated analysis and representation of 3D polycrystalline microstructures, part 1: statistical characterization. Acta Mat 56(6):1257–1273

Groeber MA, Ghosh S, Uchic MD, Dimiduk D (2008b) A framework for automated analysis and representation of 3D polycrystalline microstructures, part 2: synthetic structure generation. Acta Mat 56(6):1274–1287

Guo EY, Chawla N, Jing T, Torquato S, Jiao Y (2014) Accurate modeling and reconstruction of three-dimensional percolating filamentary microstructures from two-dimensional micrographs via dilation-erosion method. Mat Charac 89:33–42

Hasanabadi A, Baniassadi M, Abrinia K, Safdari M, Garmestani H (2016) 3D microstructural reconstruction of heterogeneous materials from 2D cross sections: a modified phase-recovery algorithm. Comput Mat Sci 111:107–115

Hill R (1963) Elastic properties of reinforced solids: some theoretical principles. J Mech Phys Sol 11(5):357–372

Jackson M (2018) DREAM.3D 6.4 Release. http://dream3d.bluequartz.net/?page_id=32

Jiao Y, Stillinger FH, Torquato S (2007) Modeling heterogeneous materials via two-point correlation functions: basic principles. Phys Rev E 76(3):031110

Jiao Y, Padilla E, Chawla N (2013) Modeling and predicting microstructure evolution in lead-/tin alloy via correlation functions and stochastic material reconstruction. Acta Mat 61(9): 3370–3377

Keshavarz S, Ghosh S (2015) Hierarchical crystal plasticity fe model for nickel-based superalloys: sub-grain microstructures to polycrystalline aggregates. Int J Solids Struct 55:17–31

Kumar A, Nguyen L, DeGraef M, Sundararaghavan V (2016) A Markov random field approach for microstructure synthesis. Model Simul Mater Sci Eng 24(3):035015

Lenthe W (2017) Twin related domains in polycrystalline nickel-base superalloys: 3d structure and fatigue. PhD thesis, University of California- Santa Barbara

Li CH, Lee CK (1993) Minimum cross entropy thresholding. Pattern Recogn 26(4):617–625

MacSleyne J, Uchic MD, Simmons JP, De Graef M (2009) Three-dimensional analysis of secondary $\gamma$' precipitates in rené-88 DT and UMF-20 superalloys. Acta Mat 57(20): 6251–6267

Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 46(253): 68–78

McDowell D, Ghosh S, Kalidindi S (2011) Representation and computational structure-property relations of random media. JOM J Miner Met Mater Soc 63(3):45–51

Meyer F (1994) Topographic distance and watershed lines. Signal Process 38(1):113–125

Niezgoda S, Turner D, Fullwood D, Kalidindi S (2010) Optimized structure based representative volume element sets reflecting the ensemble-averaged 2-point statistics. Acta Mat 58:4432–4445

Nouailhas D, Cailletaud G (1996) Multiaxial behaviour of Ni-base single crystals. Scrip Mat 34:565–571

Ostoja-Starzewski M (2006) Material spatial randomness: from statistical to representative volume element. Prob Eng Mech 21(2):112–132

Parthasarathy TA, Rao SI, Dimiduk D (2004) A fast spreadsheet model for the yield strength of superalloys. In: Green KA, Pollock TM, Harada H, Howson TE, Reed RC, Schirra JJ, Walston S (eds) TMS (The Minerals, Metals & Materials Society), Superalloys, pp 887–896

Pilchak A (2013) Fatigue crack growth rates in alpha titanium: faceted vs. striation growth. Scrip Mat 68(5):277–280

Pinz M, Weber G, Lenthe W, Uchic M, Pollock T, Woodward C, Ghosh S (2018) Microstructure and property based statistically equivalent representative volume elements for modeling sub-grain $\gamma - \gamma'$ microstructures in Ni-based superalloys. Acta Mater 157:245–258

Pollock TM, Argon A (1992) Creep resistance of CMSX-3 Nickel-base superalloy single crystals. Acta Mat 40:1–30

Pollock TM, Tin S (2006) Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. J Propuls Power 22(2):361–374

Pyrz R (2006) Quantitative description of the microstructure of composites. part I: morphology of unidirectional composite systems. Comput Sci Technol 50(2):197–208

Rollett AD, Robert C, Saylor D (2006) Three dimensional microstructures: statistical analysis of second phase particles in AA7075-T651. Mater Sci Forum 519–521:1–10

Rollett AD, Lee SB, Campman R, Rohrer GS (2007) Three-dimensional characterization of microstructure by electron back-scatter diffraction. Annu Rev Mater Res 37:627–658

Saylor DM, Fridy J, El-Dasher BS, Jung KY, Rollett AD (2004) Statistically representative 3D microstructures based on orthogonal observation sections. Metall Mater Trans A 35:1969–1979

Sundararaghavan V, Zabaras N (2005) Classification and reconstruction of three-dimensional microstructures using support vector machines. Comput Mater Sci 32(2):223–239

Swaminathan S, Ghosh S (2006) Statistically equivalent representative volume elements for composite microstructures, part I: with interfacial debonding. J Compos Mater 40(7):605–621

Swaminathan S, Ghosh S, Pagano NJ (2006) Statistically equivalent representative volume elements for composite microstructures, part I: without damage. J Compos Mater 40(7): 583–604

Tewari A, Gokhale AM, Spowart JE, Miracle DB (2004) Quantitative characterization of spatial clustering in three-dimensional microstructures using two-point correlation functions. Acta Mater 52(2):307–319

Torquato S (2002) Random heterogeneous materials. Springer, New York

Tu X, Shahba A, Ghosh S (2019) Microstructure and response-based statistically equivalent RVEs for 7000-series aluminum alloys. Int J Plast 115: 268–292. https://doi.org/10.1016/j.ijplas.2018.12.002

Turner TJ, Shade PA, Bernier JV, Li SF, Schuren JC, Kenesei P, Suter RM, Almer J (2017) Crystal plasticity model validation using combined high-energy diffraction microscopy data for a Ti-7Al specimen. Metall Mater Trans A 48:627–647

Uchic MD, Groeber MA, Dimiduk DM, Simmons JP (2006) 3D microstructural characterization of nickel superalloys via serial-sectioning using a dual beam FIB-SEM. Scripta Mat 55(1):23–28

Underwood E (1972) The mathematical foundations of quantitative stereology. In: Stereology and quantitative metallography. American Society for Testing and Materials, Philadelphia, pp 3–38

Unocic RR, Zhou N, Kovarik L, Shen C, Wang Y, Mills MJ (2011) Dislocation decorrelation and relationship to deformation microtwins during creep of a $\gamma'$ precipitate strengthened Ni-based superalloy. Acta Mater 59:7325–7339

Wang Q, Zhang H, Cai H, Fan Q, Zhang X (2016) Statistical three-dimensional reconstruction of co-continuous ceramic composites. Finite Elem Anal Des 114:85–91

Zhang C, Enomoto M, Suzuki A, Ishimaru T (2004) Characterization of three-dimensional grain structure in polycrystalline iron by serial sectioning. Metall Mater Trans A 35(7):1927–1933

# Polycrystal Plasticity Models Based on Green's Functions: Mean-Field Self-Consistent and Full-Field Fast Fourier Transform Formulations

**77**

Ricardo A. Lebensohn

## Contents

### Abstract

In this chapter, we review two crystal plasticity methodologies based on Green's functions for the prediction of microstructure-property relations in polycrystalline aggregates. The first, known as the viscoplastic self-consistent (VPSC) formulation, is a mean-field theory. The second, known as the viscoplastic fast Fourier transform-based (VPFFT) formulation, is a full-field method. The assumptions and main equations of these Green's function-based crystal plasticity formulations are presented using a unified notation, pointing out their similarities and differences and cross-validating their predictions.

R. A. Lebensohn (✉)
Fluid Dynamics and Solid Mechanics, T-3 Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA
e-mail: lebenso@lanl.gov

# 1    Introduction

An accurate prediction of the mechanical behavior of polycrystalline aggregates undergoing plastic deformation based on their evolving microstructure and the directional properties and substructure of their constituent single crystal grains is indispensable to establish physically based relationships between microstructure and properties of this ubiquitous class of materials.

On the one hand, advances in the theories linking microstructures and properties of heterogeneous materials, in particular approaches based on Green's functions, have enabled the development of very accurate mean-field models for the prediction of the effective plastic response of statistically defined classes of polycrystalline aggregates. On the other hand, novel and very efficient full-field approaches also based on Green's function formalism have been proposed and applied to the prediction of the actual micromechanical fields developing inside the grains of polycrystalline unit cells with specific microstructures. In this chapter, the assumptions and main equations of these Green's function-based crystal plasticity formulations are presented using a unified notation, pointing out their similarities and differences and cross-validating their predictions.

Concerning mean-field approximations, the computation of effective mechanical response and texture evolution of polycrystalline materials using homogenization approaches has a long tradition (e.g., Sachs 1928; Taylor 1938). Presently, self-consistent approximations are extensively used to deal with this problem. The one-site viscoplastic (VP) self-consistent (SC) theory of polycrystal deformation can be traced back to the works of Molinari et al. (1987), who established an homogenization procedure based on an iterative method involving the computation of integrals in ellipsoidal domains of the *infinite medium* Green's function, customarily used in the solution of the partial derivative equations (PDEs) governing the micromechanical response of heterogeneous materials, and Lebensohn and Tomé (1993), who implemented numerically this formulation to fully account for polycrystal anisotropy, developing the first version of the VPSC code. Since its inception, the VPSC code has experienced several improvements and extensions, e.g., Wenk et al. 1997 (recrystallization), Lebensohn and Canova 1997 (two-site VPSC approximation for two-phase polycrystals), Lebensohn et al. 1998 (VPSC modeling of lamellar structures), Lebensohn et al. 2007 (second-order linearization), Proust et al. 2007 (improved VPSC modeling of twinning), Beyerlein and Tomé 2008 (dislocation density-based hardening models), Lebensohn et al. 2011 (*dilatational* VPSC for porous polycrystals), Lebensohn et al. 2016 (lattice rotation rate fluctuation calculation), Wen et al. 2016 (improved hardening laws for strain path changes), Zecevic et al. 2017 (VPSC prediction of intragranular misorientation evolution), etc., and it is nowadays extensively used to simulate plastic deformation of polycrystalline aggregates and for interpretation of experimental evidence on metals, minerals, and polymers. (Access to the VPSC code can be obtained from the author by email request.) Moreover, beyond applications involving single material point calculations, VPSC has been adapted to account for microstructural effects in multiscale calculations by either coupling it directly with finite elements (FE)

(e.g., Tomé et al. 2001; Segurado et al. 2012; Knezevic et al. 2013; Barton et al. 2015) or using VPSC in combination with uniaxial experiments to fit anisotropic yield functions for subsequent use in the FE analysis (e.g., Plunkett et al. 2006; Nixon et al. 2010; Knezevic et al. 2013). These new multiscale developments greatly expanded the scope of applications of the model.

The self-consistent theory is one of the most commonly used homogenization methods to estimate the mechanical response behavior of polycrystals and was originally proposed by Hershey (1954) for linear elastic materials. For nonlinear aggregates (as those formed by grains deforming in the VP regime), several SC approximations that were subsequently proposed differ in the procedure used to linearize the nonlinear mechanical behavior at grain level, but eventually all of them end up making use of the original linear SC theory. Among the nonlinear SC formulations, we can mention the secant (Hill 1965; Hutchinson 1976), the tangent (Molinari et al. 1987; Lebensohn and Tomé 1993), and the affine (Masson et al. 2000) approximations. All these are first-order SC approximations, since they are based on linearization schemes that, at grain level, make use of information on field averages only, disregarding higher-order statistical information inside the grains. However, the above assumption may be questionable, especially for materials having strong directionality and/or large variations in local properties, as in the case of low rate-sensitivity materials, aggregates made of highly anisotropic grains, voided and/or multiphase polycrystals. In all those cases, strong deformation gradients are likely to develop inside grains because of differences in properties with neighboring crystals.

To overcome the above limitations, more accurate nonlinear homogenization methods were developed by Ponte Castañeda and collaborators, using linearization schemes at grain level that also incorporate accessible information on the second moments of the stress field distributions in the grains. These more elaborate SC formulations are based on the concept of linear comparison material, which express the effective potential of the nonlinear VP polycrystal in terms of that of a linearly viscous aggregate with properties that are determined from suitably designed variational principles. Among these improved linearization schemes, the second-order method, proposed for nonlinear composites (Ponte Castañeda 2002), later extended to VP polycrystals (Liu and Ponte Castañeda 2004) and implemented in the VPSC code (Lebensohn et al. 2007), uses the SC approximation for a general class of linearly viscous polycrystals, having a nonvanishing strain rate at zero stress, to generate the most accurate SC estimates for VP polycrystals.

Until recently, another limitation of the VPSC formulation was that only the average values of the micromechanical fields inside the single crystal grains were used to update the microstructure as deformation proceeded. In particular, the average lattice rotation calculated for each grain was applied to update its average orientation after each deformation increment, and therefore the evolution of deformation texture was based on these average orientation updates. This approximation, which neglects the progressive buildup of intragranular misorientation, has two deleterious effects on the quality of the predicted texture and microstructure evolution. First, the predicted deformation textures are systematically sharper than the experimental

textures. Second, since intragranular misorientation is not accounted for, phenomena like grain size reduction by grain fragmentation and other important mechanisms affecting microstructure evolution that are driven in part by local lattice misorientation, like recovery and recrystallization, cannot be predicted based on strictly micromechanical considerations. Below it is reported how the numerically tractable problem of estimating average fluctuations of the stress field distribution inside each grain – already implemented in the VPSC code as part of the second-order linearization procedure – can also be used to calculate the corresponding second moments of the lattice rotation rate field in each grain.

In what concerns full-field approaches, crystal plasticity finite element (CPFE) implementations have been extensively applied to obtain solutions for the plastic deformation of polycrystalline materials with intracrystalline resolution (e.g., Becker 1991; Mika and Dawson 1998; Delaire et al. 2001; Barbe et al. 2001; Raabe et al. 2001; Delannay et al. 2006; Musienko et al. 2007; Roters et al. 2010; Cruzado et al. 2015; Cheng and Ghosh 2015; Quey et al. 2015, etc.). However, the large number of degrees of freedom required by such FE calculations limits the size of the microstructures that can be simulated by these methods. Conceived as a very efficient alternative to FE methods, a numerical formulation based on the use of the efficient fast Fourier transform (FFT) algorithm to compute convolution integrals over the entire periodic unit cell involving the *periodic* Green's function was originally proposed by Moulinec and Suquet (1994) for the prediction of the micromechanical behavior of heterogeneous materials. The latter includes both composites (e.g., Moulinec and Suquet 1998; Eyre and Milton 1999; Michel et al. 2000; Idiart et al. 2006; Brisard and Dormieux 2010; Willot 2015, etc.), in which the source of heterogeneity is related to the spatial distribution of phases with different mechanical properties, and polycrystals (e.g., Lebensohn 2001; Lebensohn et al. 2005, 2008, 2012, 2013; Brenner et al. 2009; Grennerat et al. 2012; Eisenlohr et al. 2013; Lebensohn and Needleman 2016, etc.), in which the heterogeneity is related to the spatial distribution of crystals with directional mechanical properties.

## 2    Models

### 2.1    VPSC Model

The self-consistent formulation consists in representing a polycrystal by means of weighted, ellipsoidal, statistically representative (SR) grains. Each of these SR grains represents the average behavior of all the grains with a particular crystallographic orientation and morphology but different environments. These SR grains should be regarded as representing the behavior of mechanical phases, i.e., all the single crystals with a given orientation (r) belong to mechanical phase (r) and are represented by SR grain (r). Note the difference between *mechanical phases*, which differ from each other only in terms of crystallographic orientation and/or morphology, and actual *phases* differing from each other in crystallographic structure and/or composition. In what follows, *SR grain (r)* and *mechanical phase*

*(r)* will be used interchangeably. Each representative grain will be treated as an ellipsoidal viscoplastic inclusion embedded in an effective viscoplastic medium. Plastic deformation in the inclusion is accommodated by dislocation slip activated by a resolved shear stress.

### 2.1.1 Local Constitutive Behavior and Homogenization

Let us consider a macroscopic velocity gradient $V_{i,j}$ applied to a polycrystalline aggregate, which can be decomposed into an average symmetric strain rate $\dot{E}_{ij} = \frac{1}{2}\left(V_{i,j} + V_{j,i}\right)$ and an average antisymmetric rotation rate $\dot{\Omega}_{ij} = \frac{1}{2}\left(V_{i,j} - V_{j,i}\right)$ and that the plastic component of the deformation is much larger than the elastic part such that the latter can be neglected and the flow is incompressible. Under these assumptions, we can use small-strain kinematics to express the constitutive behavior as a relation between Cauchy stress and velocity gradient in the current configuration. Once the local velocity gradient field is obtained, the evolution of microstructure and micromechanical variables can be calculated by integrating the velocity gradient field in small time increments to explicitly update the current configuration of the material. Under this kinematic framework, the viscoplastic constitutive behavior at each material point $\mathbf{x}$ can be phenomenologically described by means of the following nonlinear, *rate-sensitive equation*:

$$
\begin{aligned}
\dot{\boldsymbol{\varepsilon}}(\mathbf{x}) = \sum_k \mathbf{m}^k(\mathbf{x}) \;\; \dot{\gamma}^k(\mathbf{x}) &= \dot{\gamma}_o \sum_k \mathbf{m}^k(\mathbf{x}) \left( \frac{\left| \mathbf{m}^k(\mathbf{x}) : \boldsymbol{\sigma}'(\mathbf{x}) \right|}{\tau_o^k(\mathbf{x})} \right)^n \\
&\times \operatorname{sgn}\left( \mathbf{m}^k(\mathbf{x}) : \boldsymbol{\sigma}'(\mathbf{x}) \right)
\end{aligned}
\tag{1}
$$

where $\tau_o^k$ is the threshold resolved shear stress of slip system (k); $m_{ij}^k = \frac{1}{2}\left(n_i^k b_j^k + n_j^k b_i^k\right)$ is the symmetric Schmid tensor associated with slip system (k), where $\mathbf{n}^k$ and $\mathbf{b}^k$ are the normal and Burgers vector direction of such slip system; and $\dot{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\sigma}'$ are the deviatoric strain rate and stress. The local shear rate on slip system (k), $\dot{\gamma}^k$, is given by:

$$
\begin{aligned}
\dot{\gamma}^k(\mathbf{x}) = \dot{\gamma}_o \left( \frac{\left| \tau^k(\mathbf{x}) \right|}{\tau_o^k(\mathbf{x})} \right)^n \times \operatorname{sgn}\left( \tau^k(\mathbf{x}) \right) &= \dot{\gamma}_o \left( \frac{\left| \mathbf{m}^k(\mathbf{x}) : \boldsymbol{\sigma}'(\mathbf{x}) \right|}{\tau_o^k(\mathbf{x})} \right)^n \\
&\times \operatorname{sgn}\left( \mathbf{m}^k(\mathbf{x}) : \boldsymbol{\sigma}'(\mathbf{x}) \right)
\end{aligned}
\tag{2}
$$

where $\tau^k$ is the resolved shear stress on slip system (k), $\dot{\gamma}_o$ is a normalization factor, and the stress exponent $n$ is the inverse of the rate-sensitivity. If the shear rates are known, the lattice rotation rate or plastic spin associated with slip activity at single crystal material point $\mathbf{x}$ is given by:

$$-\dot{\boldsymbol{\omega}}^p(\mathbf{x}) = -\sum_k \boldsymbol{\alpha}^k(\mathbf{x}) \ \dot{\gamma}^k(\mathbf{x}) \tag{3}$$

where $\alpha_{ij}^k = \frac{1}{2}\left(n_i^k b_j^k - n_j^k b_i^k\right)$ is the antisymmetric Schmid tensor. Note that although Eqs. (1), (2), and (3) can be used to deal with crystal deforming by slip and twinning, in the examples that follows (in the context of both homogenization and full-field approaches), we will only consider crystal deformation by slip. In this way we avoid the additional complication of having to deal with twinning reorientation.

Let us assume that the following linear relations, which are approximations of the actual nonlinear relations, Eqs. (1) and (2), hold for the SR grain (r), i.e., for $\mathbf{x} \in \Omega^{(r)}$:

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{x}) = \mathbf{M}^{(r)} : \boldsymbol{\sigma}'(\mathbf{x}) + \dot{\boldsymbol{\varepsilon}}^{o\,(r)} \tag{4}$$

$$\dot{\gamma}^k(\mathbf{x}) = \eta^{k(r)} \ \tau^k(\mathbf{x}) + \dot{g}^{ok(r)} \tag{5}$$

where the moduli $\mathbf{M}^{(r)}$ and $\dot{\boldsymbol{\varepsilon}}^{o\,(r)}$ are the compliance and back-extrapolated strain rate (strain rate under zero stress) of grain (r), respectively, and $\eta^{k(r)}$ and $\dot{g}^{ok(r)}$ are the slip-level compliance and back-extrapolated shear rate, respectively. Depending on the linearization assumption, the above moduli can be chosen differently. For example, under the first-order affine linearization, the moduli are given by:

$$\mathbf{M}^{(r)} = n\dot{\gamma}_o \sum_k \frac{\mathbf{m}^{k(r)} \otimes \ \mathbf{m}^{k(r)}}{\tau_o^{k(r)}} \left(\frac{\mathbf{m}^{k(r)} : \ \boldsymbol{\sigma}'^{(r)}}{\tau_o^{k(r)}}\right)^{n-1} \text{and} \quad \dot{\boldsymbol{\varepsilon}}^{o(r)} = (1-n) \ \dot{\boldsymbol{\varepsilon}}^{(r)} \tag{6}$$

$$\eta^{k(r)} = n\frac{\dot{\gamma}_o}{\tau_o^{k(r)}} \left(\frac{\tau^{k(r)}}{\tau_o^{k(r)}}\right)^{n-1} \text{and} \quad \dot{g}^{ok(r)} = (1-n) \ \dot{\gamma}^{k(r)} \tag{7}$$

where index (r) indicates average (first moment) of the field over SR grain (r). Following Hill (1965) and Hutchinson (1976), let us express the homogenized behavior of a linear heterogeneous medium whose local behavior is described by Eq. (4) assuming an analogous linear relation at the effective medium (macroscopic) level:

$$\dot{\mathbf{E}} = \overline{\mathbf{M}} : \boldsymbol{\Sigma}' + \dot{\mathbf{E}}^o \tag{8}$$

where $\dot{\mathbf{E}}$ and $\boldsymbol{\Sigma}'$ are the overall (macroscopic) deviatoric strain rate and stress tensors and $\overline{\mathbf{M}}$ and $\dot{\mathbf{E}}^o$ are, respectively, the viscoplastic compliance and back-extrapolated term of an a priori unknown homogeneous equivalent medium (HEM). The procedure followed to obtain this homogenized response is the linear self-consistent method. The problem underlying the self-consistent method is that of an

inhomogeneous domain (r) of moduli $\mathbf{M}^{(r)}$ and $\dot{\boldsymbol{\varepsilon}}^{o(r)}$, embedded in an infinite medium of moduli $\overline{\mathbf{M}}$ and $\dot{\mathbf{E}}^o$. Invoking the concept of the equivalent inclusion (Mura 1987), the local constitutive behavior in domain (r) can be rewritten as:

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{x}) = \overline{\mathbf{M}} : \boldsymbol{\sigma}'(\mathbf{x}) + \dot{\mathbf{E}}^o + \dot{\boldsymbol{\varepsilon}}^*(\mathbf{x}) \tag{9}$$

where $\dot{\boldsymbol{\varepsilon}}^*(\mathbf{x})$ is an eigen-strain-rate field, which follows from replacing the inhomogeneity by an equivalent inclusion. Rearranging and subtracting Eq. (8) from Eq. (9) gives:

$$\tilde{\boldsymbol{\sigma}}'(\mathbf{x}) = \overline{\mathbf{L}} : \left( \dot{\tilde{\boldsymbol{\varepsilon}}}(\mathbf{x}) - \dot{\boldsymbol{\varepsilon}}^*(\mathbf{x}) \right) \tag{10}$$

The symbol "$\sim$" denotes local deviations from macroscopic values of the corresponding magnitudes, and $\overline{\mathbf{L}} = \overline{\mathbf{M}}^{-1}$. Combining Eq. (10) with the equilibrium condition gives:

$$\sigma_{ij,j}(\mathbf{x}) = \tilde{\sigma}_{ij,j}(\mathbf{x}) = \tilde{\sigma}'_{ij,j}(\mathbf{x}) + \tilde{\sigma}^m_{,i}(\mathbf{x}) \tag{11}$$

where $\sigma_{ij}$ and $\sigma^m$ are the Cauchy stress tensor and the mean stress, respectively. Using the relation $\dot{\tilde{\varepsilon}}_{ij}(\mathbf{x}) = \frac{1}{2}\left(\tilde{v}_{i,j}(\mathbf{x}) + \tilde{v}_{j,i}(\mathbf{x})\right)$ between the strain-rate and velocity gradient deviations, and adding the incompressibility condition associated with plastic deformation, we obtain:

$$\left| \begin{array}{l} \overline{L}_{ijkl} \ \tilde{v}_{k,lj}(\mathbf{x}) + \tilde{\sigma}^m_{,i}(\mathbf{x}) + \phi_{ij,j}(\mathbf{x}) = 0 \\ \tilde{v}_{k,k}(\mathbf{x}) = 0 \end{array} \right. \tag{12}$$

where:

$$\varphi_{ij}(\mathbf{x}) = -\overline{L}_{ijkl}\dot{\varepsilon}^*_{kl}(\mathbf{x}) \tag{13}$$

is a heterogeneity or *polarization field*, and its divergence $f_i(\mathbf{x}) = \varphi_{ij,j}(\mathbf{x})$ is an (artificial) external volumetric force field applied to the material. System (12) consists of four differential equations with four unknowns: three are the components of velocity deviation vector $\tilde{v}_i(\mathbf{x})$, and one is the mean stress deviation $\tilde{\sigma}^m(\mathbf{x})$. A system of N linear differential equations with N unknown functions and a polarization term can be solved using Green's function method. Let us call $G_{km}(\mathbf{x})$ and $H_m(\mathbf{x})$ Green's functions associated with $\tilde{v}_i(\mathbf{x})$ and $\tilde{\sigma}^m(\mathbf{x})$, respectively, which solve the auxiliary problem of a unitary volumetric force, with a single nonvanishing m-component:

$$\left| \begin{array}{l} \overline{L}_{ijkl} \ G_{km,lj}\left(\mathbf{x} - \mathbf{x}'\right) + H_{m,i}\left(\mathbf{x} - \mathbf{x}'\right) + \delta_{im}\,\delta\left(\mathbf{x} - \mathbf{x}'\right) = 0 \\ G_{km,k}\left(\mathbf{x} - \mathbf{x}'\right) = 0 \end{array} \right. \tag{14}$$
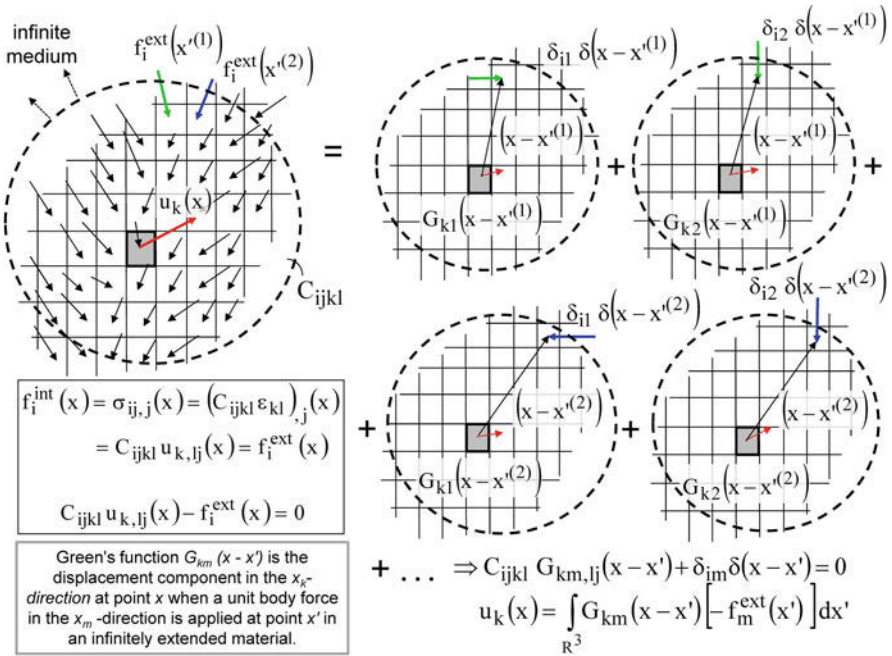
**Fig. 1** Schematic illustration of Green's function associated with the displacement field of an elastic infinite medium. $G_{km}(\mathbf{x} - \mathbf{x}')$ is the component of the displacement in the $x_k$-direction at point $\mathbf{x}$ when a unit external force in the $x_m$-direction is applied at point $\mathbf{x}'$. Red arrows indicate displacement vectors at point $\mathbf{x}$ in reaction to external forces applied at different points $\mathbf{x}'$ indicated by black, green, and blue arrows

Green's function $G_{km}(\mathbf{x} - \mathbf{x}')$ is the local deviation in displacement (or velocity) component in the $x_k$-direction at point $\mathbf{x}$ when a unit body force in the $x_m$-direction is applied at point $\mathbf{x}'$ in an infinitely extended material. In the incompressible case of Eq. (14), $H_m(\mathbf{x} - \mathbf{x}')$ is the corresponding response in mean stress. For illustration, Fig. 1 shows a schematic explanation of Green's function of the displacement field $G_{km}(\mathbf{x} - \mathbf{x}')$ for an infinite (compressible) elastic medium.

Once the solution of Eq. (14) is obtained, the solution for the velocity field is given by the convolution integral:

$$\tilde{v}_k(\mathbf{x}) = \int_{R^3} G_{ki}(\mathbf{x} - \mathbf{x}') \; f_i(\mathbf{x}') \; d\mathbf{x}' = \int_{R^3} G_{ki}(\mathbf{x} - \mathbf{x}') \; \varphi_{ij,j}(\mathbf{x}') \; d\mathbf{x}' \qquad (15)$$

System (14) can be solved using Fourier transforms. Expressing Green's functions in terms of their inverse Fourier transforms, the differential system (14) can be transformed into an algebraic system:

$$\left|\begin{matrix} \alpha_j \ \alpha_l \ \overline{L}_{ijkl} \ k^2 \hat{G}_{km}(\boldsymbol{\xi}) + \alpha_i \ ik\hat{H}_m(\boldsymbol{\xi}) = \delta_{im} \\ \alpha_k \ k^2 \hat{G}_{km}(\boldsymbol{\xi}) = 0 \end{matrix}\right. \tag{16}$$

where k and $\boldsymbol{\alpha}$ are the modulus and the unit vector associated with a point of Fourier space $\boldsymbol{\xi} = k\boldsymbol{\alpha}$, respectively. Solving system (16) (see Lebensohn et al. 2007) one obtains:

$$k^2 \hat{G}_{ij}(\boldsymbol{\xi}) = A_{ij}^{-1}(\boldsymbol{\alpha}) \tag{17}$$

where $A_{ik}'(\boldsymbol{\alpha}) = \alpha_j \ \alpha_l \ \overline{L}_{ijkl}$ and

$$A(\boldsymbol{\alpha}) \quad = \quad \begin{vmatrix} A_{11}' & A_{12}' & A_{13}' & \alpha_1 \\ A_{21}' & A_{22}' & A_{23}' & \alpha_2 \\ A_{31}' & A_{32}' & A_{33}' & \alpha_3 \\ \alpha_1 & \alpha_2 & \alpha_3 & 0 \end{vmatrix} \tag{18}$$

Knowing Green's function expression in Fourier space, we can write the solution of the eigen-strain-rate problem using convolution integrals. Taking partial derivatives to Eq. (15) and after some manipulation (see Lebensohn et al. 2007; Mura 1987), we obtain:

$$\tilde{v}_{k,l}(\mathbf{x}) = \int_{R^3} G_{ki,jl}(\mathbf{x} - \mathbf{x}') \ \varphi_{ij}(\mathbf{x}') \ d\mathbf{x}' \tag{19}$$

The integral Eq. (19) provides an exact implicit solution to the problem. Furthermore, it is known from Eshelby's elastic inclusion formalism that if the eigen-strain is uniform over an ellipsoidal domain where the stiffness tensor is uniform, then the stress and strain are constant over the domain of the inclusion (r). The latter suggests the use of an a priori unknown constant polarization within the volume $\Omega$ of the ellipsoidal inclusion. This allows us to average the local field (19) over the domain $\Omega$ and obtain an average strain rate inside the inclusion of the form:

$$\tilde{v}_{k,l}^{(r)} = T_{klij} \ \overline{L}_{ijmn} \ \dot{\varepsilon}_{mn}^{*(r)} \tag{20}$$

where $T_{klij}$ is an integral of Green's function over domain $\Omega$. For an ellipsoidal inclusion of radii (a,b,c) (see Berveiller et al. 1987; Lebensohn et al. 2007 for details), $T_{klij}$ can be reduced to:

$$T_{klij} = \frac{abc}{4\pi} \int_0^{2\pi} \int_0^{\pi} \frac{\alpha_j \ \alpha_l \ A_{ki}^{-1}(\boldsymbol{\alpha})}{[\rho(\boldsymbol{\alpha})]^3} \sin\theta \ d\theta \ d\varphi \tag{21}$$

where $\rho(\boldsymbol{\alpha}) = [\ (a\alpha_1)^2 + (b\alpha_2)^2 + (c\alpha_3)^2]^{1/2}$. The symmetric and antisymmetric Eshelby tensors (functions of $\overline{L}$ and the shape of the ellipsoidal inclusion, representing the morphology of the SR grains) are defined as:

$$S_{ijkl} = \frac{1}{4}\left(T_{ijmn} + T_{jimn} + T_{ijnm} + T_{jinm}\right)\ \overline{L}_{mnkl} \tag{22a}$$

$$\Pi_{ijkl} = \frac{1}{4}\left(T_{ijmn} - T_{jimn} + T_{ijnm} - T_{jinm}\right)\ \overline{L}_{mnkl} \tag{22b}$$

Taking symmetric and antisymmetric components to Eq. (19) and using Eqs. (22), we obtain the average strain-rate and rotation rate deviations in the ellipsoidal domain:

$$\tilde{\dot{\boldsymbol{\varepsilon}}}^{(r)} = \mathbf{S} : \dot{\boldsymbol{\varepsilon}}^{*(r)} \tag{23}$$

$$\tilde{\dot{\boldsymbol{\omega}}}^{(r)} = \boldsymbol{\Pi} : \dot{\boldsymbol{\varepsilon}}^{*(r)} = \boldsymbol{\Pi} : \mathbf{S}^{-1} : \tilde{\dot{\boldsymbol{\varepsilon}}}^{(r)} \tag{24}$$

where $\tilde{\dot{\boldsymbol{\varepsilon}}}^{(r)} = \dot{\mathbf{E}} - \dot{\boldsymbol{\varepsilon}}^{(r)}$ and $\tilde{\dot{\boldsymbol{\omega}}}^{(r)} = \dot{\boldsymbol{\Omega}} - \dot{\boldsymbol{\omega}}^{(r)}$ are deviations of the average strain rate and rotation rate inside the inclusion, with respect to the corresponding overall magnitudes, and $\dot{\boldsymbol{\varepsilon}}^{*(r)}$ is the average eigen-strain rate in the inclusion. Therefore, the lattice rotation rate field is given by:

$$\dot{\boldsymbol{\omega}}\ (\mathbf{x}) = \dot{\boldsymbol{\Omega}} + \tilde{\dot{\boldsymbol{\omega}}}^{(r)} - \dot{\boldsymbol{\omega}}^p\ (\mathbf{x}) \tag{25}$$

### 2.1.2   Interaction and Localization Equations

Taking volume averages over the domain of the inclusion on both sides of Eq. (10) gives:

$$\tilde{\boldsymbol{\sigma}}'^{(r)} = \overline{\mathbf{L}} : \left(\tilde{\dot{\boldsymbol{\varepsilon}}}^{(r)} - \dot{\boldsymbol{\varepsilon}}^{*(r)}\right) \tag{26}$$

Replacing the eigen-strain rate given by Eq. (23) into Eq. (26), we obtain the *interaction equation*:

$$\tilde{\dot{\boldsymbol{\varepsilon}}}^{(r)} = -\tilde{\mathbf{M}} : \tilde{\boldsymbol{\sigma}}'^{(r)} \tag{27}$$

where the interaction tensor is given by:

$$\tilde{\mathbf{M}} = (\mathbf{I} - \mathbf{S})^{-1} : \mathbf{S} : \overline{\mathbf{M}} \tag{28}$$

Replacing the constitutive relations of the inclusion and the effective medium in the interaction equation and after some manipulation, one can write the following *localization equation*:

$$\boldsymbol{\sigma}'^{(r)} = \mathbf{B}^{(r)} : \boldsymbol{\Sigma}' + \mathbf{b}^{(r)} \tag{29}$$

where the localization tensors are defined as:

$$\mathbf{B}^{(r)} = \left(\mathbf{M}^{(r)} + \tilde{\mathbf{M}}\right)^{-1} : \left(\overline{\mathbf{M}} + \tilde{\mathbf{M}}\right) \quad \text{and} \quad \mathbf{b}^{(r)} = \left(\mathbf{M}^{(r)} + \tilde{\mathbf{M}}\right)^{-1} : \left(\dot{\mathbf{E}}^{o} - \dot{\boldsymbol{\varepsilon}}^{o(r)}\right) \tag{30}$$

### 2.1.3   Self-Consistent Equations

The derivation presented in the previous sections solves the problem of an equivalent inclusion embedded in an effective medium. In this section, we use the previous result to construct a polycrystal model, consisting in regarding each SR grain (r) as an inclusion embedded in an effective medium that represents the polycrystal. The properties of such medium are not known a priori but have to be found through an iterative procedure. Replacing the stress localization equation (Eq. 29) in the local constitutive equation (Eq. 4) averaged over the SR grain (r), we obtain:

$$\dot{\boldsymbol{\varepsilon}}^{(r)} = \mathbf{M}^{(r)} : \mathbf{B}^{(r)} : \boldsymbol{\Sigma} + \mathbf{M}^{(r)} : \mathbf{b}^{(r)} + \dot{\boldsymbol{\varepsilon}}^{o(r)} \tag{31}$$

Taking volumetric average to Eq. (31), enforcing the condition that the average of the strain rates over the aggregate has to coincide with the macroscopic quantities, i.e.:

$$\dot{\mathbf{E}} = \left\langle \dot{\boldsymbol{\varepsilon}}^{(r)} \right\rangle \tag{32}$$

(where the brackets "$\langle\rangle$" denote average over the SR grains, weighted by the associated volume fraction), and using the macroscopic constitutive relation (Eq. 8), we obtain the following *self-consistent equations* for the HEM's compliance and back-extrapolated term:

$$\overline{\mathbf{M}} = \left\langle \mathbf{M}^{(r)} : \mathbf{B}^{(r)} \right\rangle \tag{33}$$

$$\dot{\mathbf{E}}^{o} = \left\langle \mathbf{M}^{(r)} : \mathbf{b}^{(r)} + \dot{\boldsymbol{\varepsilon}}^{o(r)} \right\rangle \tag{34}$$

### 2.1.4   Fluctuations Calculation

The methodology to obtain average fluctuations of the stress distributions in homogeneous domains of heterogeneous media was originally derived by Bobeth and

Diener ([1987](#)), Kreher ([1990](#)), and Parton and Buryachenko ([1990](#)) for composites and extended by Lebensohn et al. ([2007](#)) for polycrystals. The effective stress potential $\overline{U}_T$ of a linearly viscous polycrystal described by Eq. ([8](#)) may be written in the form (Laws [1973](#)):

$$\overline{U}_T = \frac{1}{2}\overline{\mathbf{M}} :: \left(\mathbf{\Sigma}' \otimes \mathbf{\Sigma}'\right) + \dot{\mathbf{E}}^o : \mathbf{\Sigma}' + \frac{1}{2}\overline{G} \tag{35}$$

where $\overline{G}$ is the power under zero applied stress. Let us rewrite the self-consistent expression for $\overline{\mathbf{M}}$ and $\dot{\mathbf{E}}^o$ (Eqs. [33](#) and [34](#)) as:

$$\overline{\mathbf{M}} = \left\langle \mathbf{M}^{(r)} : \mathbf{B}^{(r)} \right\rangle = \sum_r c^{(r)} \ \mathbf{M}^{(r)} : \mathbf{B}^{(r)} \tag{36}$$

$$\dot{\mathbf{E}}^o = \left\langle \mathbf{M}^{(r)} : \mathbf{b}^{(r)} + \dot{\mathbf{\varepsilon}}^{o\ (r)} \right\rangle = \sum_r c^{(r)} \ \left( \mathbf{M}^{(r)} : \mathbf{b}^{(r)} + \dot{\mathbf{\varepsilon}}^{o\ (r)} \right) \tag{37}$$

where $c^{(r)}$ is the volume fraction associated with SR grain (r). The corresponding expression for $\overline{G}$ is:

$$\overline{G} = \sum_r c^{(r)} \ \dot{\mathbf{\varepsilon}}^{o\ (r)} : \mathbf{b}^{(r)} \tag{38}$$

The second moment of the stress field over a SR grain (r) of this polycrystal is a fourth-rank tensor given by:

$$\left\langle \mathbf{\sigma}' \otimes \mathbf{\sigma}' \right\rangle^{(r)} = \frac{2}{c^{(r)}} \frac{\partial \overline{U}_T}{\partial \mathbf{M}^{(r)}} \tag{39}$$

Replacing Eqs. ([35](#)), ([36](#)), ([37](#)), and ([38](#)) in ([39](#)), we obtain:

$$\left\langle \mathbf{\sigma}' \otimes \mathbf{\sigma}' \right\rangle^{(r)} = \frac{1}{c^{(r)}} \frac{\partial \overline{\mathbf{M}}}{\partial \mathbf{M}^{(r)}} :: \left(\mathbf{\Sigma}' \otimes \mathbf{\Sigma}'\right) + \frac{1}{c^{(r)}} \frac{\partial \dot{\mathbf{E}}^o}{\partial \mathbf{M}^{(r)}} : \mathbf{\Sigma}' + \frac{1}{c^{(r)}} \frac{\partial \overline{G}}{\partial \mathbf{M}^{(r)}} \tag{40}$$

The algorithmic expressions to calculate the partial derivatives on the right-hand side are given in Lebensohn et al. ([2007](#)). These average stress fluctuations in the grains are necessary for (a) the formulation of the more accurate second-order approximation (Liu and Ponte Castañeda [2004](#)) to linearize the behavior of the SR grains (see Lebensohn et al. [2007](#) for details of its implementation in VPSC) and (b) the calculation based on strictly micromechanical arguments of intragranular misorientations, for further improved prediction of microstructure evolution (Lebensohn et al. [2016](#)). In connection with the latter, the second moment of the plastic spin field in SR grain (r) is given by:

$$\langle \dot{\boldsymbol{\omega}}^{\mathrm{p}} \otimes \dot{\boldsymbol{\omega}}^{\mathrm{p}} \rangle^{(\mathrm{r})} = \left\langle \left( \sum_{\mathrm{k}} \boldsymbol{\alpha}^{\mathrm{k}}(\mathbf{x}) \; \dot{\gamma}^{\mathrm{k}}(\mathbf{x}) \right) \otimes \left( \sum_{\mathrm{k}} \boldsymbol{\alpha}^{\mathrm{k}}(\mathbf{x}) \; \dot{\gamma}^{\mathrm{k}}(\mathbf{x}) \right) \right\rangle^{(\mathrm{r})} \tag{41}$$

Assuming crystal orientation is uniform inside grain (r):

$$\langle \dot{\boldsymbol{\omega}}^{\mathrm{p}} \otimes \dot{\boldsymbol{\omega}}^{\mathrm{p}} \rangle^{(\mathrm{r})} = \sum_{\mathrm{k},\mathrm{k}'} \left( \boldsymbol{\alpha}^{\mathrm{k}(\mathrm{r})} \otimes \boldsymbol{\alpha}^{\mathrm{k}'(\mathrm{r})} \right) \left\langle \dot{\gamma}^{\mathrm{k}} \; \dot{\gamma}^{\mathrm{k}'} \right\rangle^{(\mathrm{r})} \tag{42}$$

Using the linearized expression for the shear rates (Eq. 5), the above can be expressed in terms of the second moment of the stress field $\langle \boldsymbol{\sigma}' \otimes \boldsymbol{\sigma}' \rangle^{(\mathrm{r})}$ (Eq. 40) as (Lebensohn et al. 2016):

$$\begin{aligned}
\langle \dot{\boldsymbol{\omega}}^{\mathrm{p}} \otimes \dot{\boldsymbol{\omega}}^{\mathrm{p}} \rangle^{(\mathrm{r})} = & \left\{ \sum_{\mathrm{k},\mathrm{k}'} \left( \eta^{\mathrm{k}(\mathrm{r})} \eta^{\mathrm{k}'(\mathrm{r})} \right) \left( \boldsymbol{\alpha}^{\mathrm{k}(\mathrm{r})} \otimes \boldsymbol{\alpha}^{\mathrm{k}'(\mathrm{r})} \right) \otimes \left( \mathbf{m}^{k(\mathrm{r})} \otimes \mathbf{m}^{\mathrm{k}'(\mathrm{r})} \right) \right\} :: \langle \boldsymbol{\sigma}' \otimes \boldsymbol{\sigma}' \rangle^{(\mathrm{r})} \\
& + \left\{ \sum_{\mathrm{k},\mathrm{k}'} \left( \boldsymbol{\alpha}^{\mathrm{k}(\mathrm{r})} \otimes \boldsymbol{\alpha}^{\mathrm{k}'(\mathrm{r})} \right) \otimes \left[ \left( \eta^{\mathrm{k}(\mathrm{r})} \dot{\mathrm{g}}^{\mathrm{ok}'(\mathrm{r})} \right) \mathbf{m}^{\mathrm{k}(\mathrm{r})} + \left( \eta^{\mathrm{k}'(\mathrm{r})} \dot{\mathrm{g}}^{\mathrm{ok}(\mathrm{r})} \right) \mathbf{m}^{\mathrm{k}'(\mathrm{r})} \right] \right\} : \langle \boldsymbol{\sigma}' \rangle^{(\mathrm{r})} \\
& + \sum_{\mathrm{k},\mathrm{k}'} \left( \boldsymbol{\alpha}^{\mathrm{k}(\mathrm{r})} \otimes \boldsymbol{\alpha}^{\mathrm{k}'(\mathrm{r})} \right) \left( \dot{\mathrm{g}}^{\mathrm{ok}(\mathrm{r})} \dot{\mathrm{g}}^{\mathrm{ok}'(\mathrm{r})} \right)
\end{aligned} \tag{43}$$

For spherical inclusions/equiaxed SR grains, $\boldsymbol{\Pi} = \mathbf{0} \Rightarrow \tilde{\dot{\boldsymbol{\omega}}}^{(\mathrm{r})} = \mathbf{0}$ (see Eq. 24), and, therefore:

$$\langle \dot{\boldsymbol{\omega}} \otimes \dot{\boldsymbol{\omega}} \rangle^{(\mathrm{r})} = \langle \dot{\boldsymbol{\omega}}^{\mathrm{p}} \otimes \dot{\boldsymbol{\omega}}^{\mathrm{p}} \rangle^{(\mathrm{r})} \tag{44}$$

The covariance of the lattice rotation rate fluctuations is obtained as:

$$\left\langle \left( \dot{\boldsymbol{\omega}} - \langle \dot{\boldsymbol{\omega}} \rangle^{(\mathrm{r})} \right) \otimes \left( \dot{\boldsymbol{\omega}} - \langle \dot{\boldsymbol{\omega}} \rangle^{(\mathrm{r})} \right) \right\rangle^{(\mathrm{r})} = \langle \dot{\boldsymbol{\omega}} \otimes \dot{\boldsymbol{\omega}} \rangle^{(\mathrm{r})} - \langle \dot{\boldsymbol{\omega}} \rangle^{(\mathrm{r})} \otimes \langle \dot{\boldsymbol{\omega}} \rangle^{(\mathrm{r})} \tag{45}$$

Representing the antisymmetric rotation rate tensors as the corresponding pseudo-vectors, the above covariance can be expressed as a symmetric matrix. Diagonalization of such covariance matrix gives an orthonormal basis of eigenvectors $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ and three associated eigenvalues: $\dot{\Omega}_1^{(\mathrm{r})} \geq \dot{\Omega}_2^{(\mathrm{r})} \geq \dot{\Omega}_3^{(\mathrm{r})}$. The square root of these eigenvalues:

$$\mathrm{SD}_{\mathrm{i}}^{(\mathrm{r})}(\dot{\boldsymbol{\omega}}) = \sqrt{\dot{\Omega}_{\mathrm{i}}^{(\mathrm{r})}} \quad (\mathrm{i} = 1, 2, 3) \tag{46}$$

are measures of the standard deviation of the rotation rate distribution along the rotation axes given by the corresponding eigenvectors. A scalar magnitude of the

rate at which intragranular orientations are spreading is given by the geometric mean of these SDs:

$$SD^{(r)}(\dot{\omega}) = \sqrt[3]{SD_1^{(r)}(\dot{\omega}) \times SD_2^{(r)}(\dot{\omega}) \times SD_3^{(r)}(\dot{\omega})} \qquad (47)$$

### 2.1.5 Numerical Implementation

To illustrate the practical use of the VPSC formulation, we describe here the steps required to predict the local and overall viscoplastic response of a polycrystal. Figure 2 shows the corresponding flow chart. Starting for convenience with an initial Taylor guess, i.e., $\dot{\varepsilon}^{(r)} = \dot{E}$, for all grains, we solve the following nonlinear equation to get $\sigma'^{(r)}$:

$$\dot{E} = \dot{\gamma}_o \sum_k \mathbf{m}^{k(r)} \left( \frac{\left| \mathbf{m}^{k(r)} : \sigma'^{(r)} \right|}{\tau_o^{k(r)}} \right)^n \times \mathrm{sgn}\left( \mathbf{m}^{k(r)} : \sigma'^{(r)} \right) \qquad (48)$$
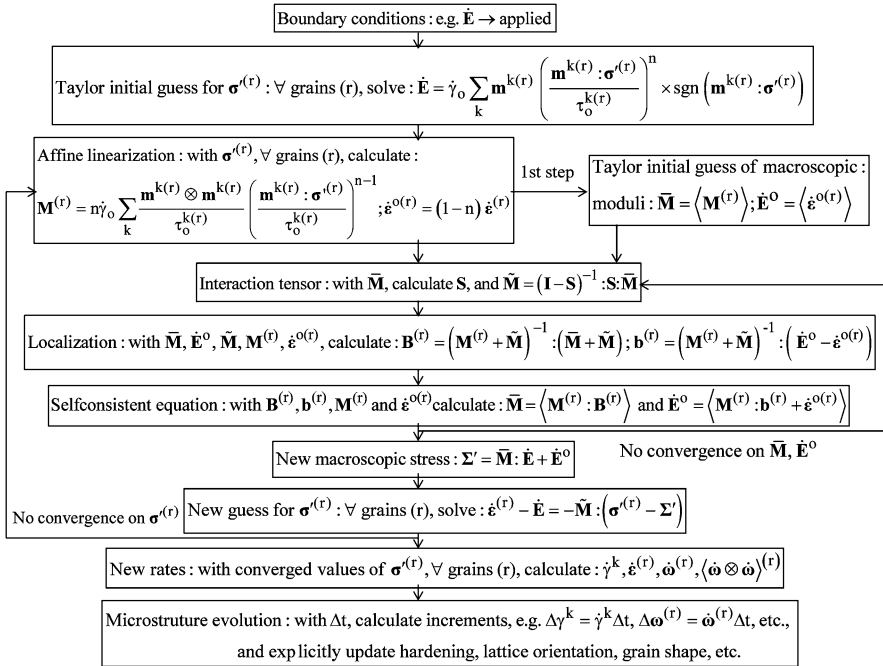


**Fig. 2** Flow chart of the numerical implementation of the VPSC formulation to predict microstructure evolution. The red arrows correspond to the two iterative loops necessary to satisfy self-consistency of the macroscopic moduli and the grain stresses

and we use an appropriate first-order linearization scheme to obtain initial values of $\mathbf{M}^{(\mathrm{r})}$ *and* $\dot{\boldsymbol{\varepsilon}}^{\mathrm{o}(\mathrm{r})}$, for each SR grain (r). Next, initial guesses for the macroscopic moduli $\overline{\mathbf{M}}$ and $\dot{\mathbf{E}}^{\mathrm{o}}$ are obtained (usually as simple averages of the local moduli). With them and the applied strain rate, the initial guess for the macroscopic stress $\boldsymbol{\Sigma}'$ can be obtained, while the Eshelby tensors $\mathbf{S}$ and $\boldsymbol{\Pi}$ can be calculated using the macroscopic moduli and the ellipsoidal shape of the SR grains, by means of the procedure described in Sect. 2.1.1. Subsequently, the interaction tensor $\tilde{\mathbf{M}}$ (Eq. 28), and the localization tensors $\mathbf{B}^{(\mathrm{r})}$ and $\mathbf{b}^{(\mathrm{r})}$ (Eq. 30), can be calculated as well. With these tensors, new estimates of $\overline{\mathbf{M}}$ and $\dot{\mathbf{E}}^{\mathrm{o}}$ are obtained, by solving iteratively the self-consistent equations (Eqs. 33 and 34). After achieving convergence on the macroscopic moduli (and, consequently, also on the macroscopic stress and the interaction and localization tensors), a new estimation of the grain stresses can be obtained, using the localization relation (Eq. 29). If the recalculated grain stresses are different (within certain tolerance) from the input values, a new iteration should be started, until convergence is reached. When the iterative procedure is completed and the grain stresses have converged, the average shear rates on each slip system are obtained using Eq. (2), and the lattice rotation rates associated with each SR grain can be calculated with the latter and Eqs. (3) and (24) and (25) averaged over the domain of each grain (r). With these rates, the above numerical scheme can be used to predict microstructure evolution, by applying viscoplastic deformation to the polycrystal in incremental steps. The latter is done explicitly, assuming constant rates during a time interval $\Delta$t (such that $\dot{\mathbf{E}}$ $\Delta t$ corresponds to a macroscopic strain increment of 1%) and using (a) lattice rotation rates (Eq. 25) (times $\Delta t$) to update crystallographic orientation of each SR grain (r) (this can be done using the average values of the latter or accounting for average rotation rate fluctuations, Zecevic et al. 2017), (b) the strain rates $\dot{\mathbf{E}} - \tilde{\dot{\boldsymbol{\varepsilon}}}^{(\mathrm{r})}$ and rigid body rotation rates $\dot{\boldsymbol{\Omega}} + \tilde{\dot{\boldsymbol{\omega}}}^{(\mathrm{r})}$ (times $\Delta t$) to update the shape and orientation of the ellipsoid representing each SR grain (r) (or, for a more efficient computation, the fluctuation terms can be neglected, keeping track of the average stretching of the grains only), and (c) the shear rates (times $\Delta t$) to update the critical stress of the deformation systems due to strain hardening, after each deformation increment.

## 2.2 VPFFT Model

The FFT-based full-field formulation for viscoplastic polycrystals is conceived for periodic unit cells, provides an "exact" solution (within the limitations imposed by the unavoidable discretization of the problem and the iterative character of the numerical algorithm, see below) of the governing equations, and has better numerical performance than a standard FE calculation for the same purpose and resolution. It was originally developed (Moulinec and Suquet 1994, 1998; Michel et al. 2000) as a fast algorithm to compute the elastic and elastoplastic effective and local response of composites and later adapted (Lebensohn 2001; Lebensohn et al. 2008) to deal with the viscoplastic deformation of polycrystals.

Briefly, the viscoplastic FFT-based formulation consists in iteratively adjusting a compatible strain-rate field, related with an equilibrated stress field through a constitutive potential, such that the average of local work-rates is minimized. The method is based on the fact that the local mechanical response of a heterogeneous medium can be calculated as a convolution integral between Green's functions associated with appropriate fields of a linear reference homogeneous medium and the actual heterogeneity field. For a periodic medium, the use can be made of the Fourier transform to reduce convolution integrals in real space to simple products in Fourier space. Thus, the FFT algorithm can be utilized to transform the heterogeneity field into Fourier space and, in turn, to get the mechanical fields by transforming that product back to real space. However, the actual heterogeneity field depends precisely on the a priori unknown mechanical fields. Therefore, an iterative scheme has to be implemented to obtain, upon convergence, a compatible strain-rate field and a stress field in equilibrium.

### 2.2.1  Periodic Unit Cell: Green Function Method

A periodic unit cell representing the polycrystal is discretized into $N_1 \times N_2 \times N_3$ Fourier points. This discretization determines a regular grid in the Cartesian space $\{\mathbf{x^d}\}$ and a corresponding grid in Fourier space $\{\boldsymbol{\xi^d}\}$. Velocities and tractions along the boundary of the unit cell are left undetermined. A velocity gradient $V_{i,j}$ (which can be decomposed into a symmetric strain rate and an antisymmetric rotation rate: $V_{i,j} = \dot{E}_{ij} + \dot{\Omega}_{ij}$) is imposed to the unit cell. The local strain-rate field is a function of the local velocity field, i.e., $\dot{\varepsilon}_{ij}(v_k(\mathbf{x}))$, and can be split into its average and a fluctuation term: $\dot{\varepsilon}_{ij}(v_k(\mathbf{x})) = \dot{E}_{ij} + \dot{\tilde{\varepsilon}}_{ij}(\tilde{v}_k(\mathbf{x}))$, where $v_i(\mathbf{x}) = \dot{E}_{ij}x_j + \tilde{v}_i(\mathbf{x})$. By imposing periodic boundary conditions, the velocity fluctuation field $\tilde{v}_k(\mathbf{x})$ is assumed to be periodic across the boundary of the unit cell, while the traction field is antiperiodic, to meet equilibrium on the boundary between contiguous unit cells.

The local constitutive relation between the strain rate $\dot{\varepsilon}_{ij}(\mathbf{x})$ and the deviatoric stress $\sigma'_{ij}(\mathbf{x})$ is given by the same rate-sensitivity relation used within the VPSC framework (Eq. 1). Let us choose a fourth-order tensor $L^o$ to be the stiffness of a linear reference medium (the choice $L^o$ of can be quite arbitrary, but the speed of convergence of the method will depend on this choice) and define the polarization field $\varphi_{ij}(\mathbf{x})$ (c.f. Eq. 13) as:

$$\varphi_{ij}(\mathbf{x}) = \tilde{\sigma}'_{ij}(\mathbf{x}) - L^o_{ijkl}\,\dot{\tilde{\varepsilon}}_{kl}(\mathbf{x}) \tag{49}$$

Then, the Cauchy stress deviation can be written as:

$$\tilde{\sigma}_{ij}(\mathbf{x}) = L^o_{ijkl}\dot{\tilde{\varepsilon}}_{kl}(\mathbf{x}) + \varphi_{ij}(\mathbf{x}) + \tilde{\sigma}^m(\mathbf{x})\,\delta_{ij} \tag{50}$$

Combining Eq. (50) with the equilibrium $\left(\sigma_{ij,j}(\mathbf{x}) = \tilde{\sigma}_{ij,j}(\mathbf{x}) = 0\right)$, the incompressibility condition, and the relation: $\dot{\tilde{\varepsilon}}_{ij}(\mathbf{x}) = \frac{1}{2}\left(\tilde{v}_{i,j}(\mathbf{x}) + \tilde{v}_{j,i}(\mathbf{x})\right)$

$$\left|\begin{array}{l} L^o_{ijkl}\tilde{v}_{k,lj}(\mathbf{x}) + \tilde{\sigma}^m_{,i}(\mathbf{x}) + \varphi_{ij,j}(\mathbf{x}) = 0 \\ \tilde{v}_{k,k}(\mathbf{x}) = 0 \end{array}\right. \tag{51}$$

This system of differential equations is formally equivalent to system (12). However, both systems actually differ in that (a) the HEM's stiffness modulus $\bar{L}$ of Eq. (12) is replaced in Eq. (51) by the stiffness of a linear reference medium $L^o$ and (b) the polarization field in Eq. (51) has in general nonvanishing values throughout the unit cell and is periodic (owing to the unit cell's periodicity), while the polarization field in Eq. (12) vanishes outside the domain of the inclusion. The auxiliary system involving periodic Green functions is then given by (c.f. Eq. 18):

$$\left|\begin{array}{l} L^o_{ijkl}\ G_{km,lj}\ (\mathbf{x} - \mathbf{x}') + H_{m,i}\ (\mathbf{x} - \mathbf{x}') + \delta_{im}\ \delta(\mathbf{x} - \mathbf{x}') = 0 \\ G_{km,k}(\mathbf{x} - \mathbf{x}') = 0 \end{array}\right. \tag{52}$$

After some manipulation, the convolution integral that gives the velocity gradient deviation field is:

$$\tilde{v}_{i,j}(\mathbf{x}) = \int_{R^3} G_{ik,jl}(\mathbf{x} - \mathbf{x}')\,\varphi_{kl}(\mathbf{x}')\,d\mathbf{x}' \tag{53}$$

Convolution integrals in direct space are simply products in Fourier space:

$$\hat{\tilde{v}}_{i,j}(\boldsymbol{\xi}) = \hat{\Gamma}_{ijkl}(\boldsymbol{\xi})\ \hat{\varphi}_{kl}(\boldsymbol{\xi}) \tag{54}$$

where the symbol "$\wedge$" indicates a Fourier transform. The Green operator in Eq. (54) is defined as $\Gamma_{ijkl} = G_{ik,jl}$. The tensors $\hat{G}_{ij}(\boldsymbol{\xi})$ and $\hat{\Gamma}_{ijkl}(\boldsymbol{\xi})$ can be calculated by taking Fourier transform to system (52):

$$\left|\begin{array}{l} \xi_l\xi_j\ L^o_{ijkl}\hat{G}_{km}(\boldsymbol{\xi}) + i\xi_i\hat{H}_m(\boldsymbol{\xi}) = \delta_{im} \\ \xi_k\hat{G}_{km}(\boldsymbol{\xi}) = 0 \end{array}\right. \tag{55}$$

Defining the $3 \times 3$ matrix $A'_{ik}(\boldsymbol{\xi}) = \xi_l\xi_j L^o_{ijkl}$, and the $4 \times 4$ matrix $A(\boldsymbol{\xi})$:

$$\mathbf{A}(\boldsymbol{\xi}) = \begin{vmatrix} A'_{11} & A'_{12} & A'_{13} & \xi_1 \\ A'_{21} & A'_{22} & A'_{23} & \xi_2 \\ A'_{31} & A'_{32} & A'_{33} & \xi_3 \\ \xi_1 & \xi_2 & \xi_3 & 0 \end{vmatrix} \tag{56}$$

we obtain from Eq. (57) (c.f. Eqs. 17–18):

$$\hat{G}_{ij}(\boldsymbol{\xi}) = A^{-1}_{ij}\ (i,j = 1,3) \quad \text{and} \quad \hat{\Gamma}_{ijkl}(\boldsymbol{\xi}) = -\xi_j\xi_l\hat{G}_{ik}(\boldsymbol{\xi}) \tag{57}$$

### 2.2.2 FFT-Based Algorithm

Assigning initial guess values to the strain-rate field in the regular grid $\{\mathbf{x^d}\}$ (e.g., $\tilde{\dot{\varepsilon}}_{ij}^{o}\left(\mathbf{x^d}\right) = 0 \Rightarrow \dot{\varepsilon}_{ij}^{o}\left(\mathbf{x^d}\right) = \dot{E}_{ij}$) and computing the corresponding stress field $\sigma_{ij}^{\prime o}\left(\mathbf{x^d}\right)$ from the local constitutive relation (Eq. 1), which requires to know the initial values of the critical stresses $\tau_{o}^{s}\left(\mathbf{x^d}\right)$ (these can be assumed, e.g., to have an uniform value $\forall\mathbf{x^d}$, adjusted to reproduce the initial macroscopic yield stress of the polycrystal) and the Schmid tensors $m_{ij}^{s}\left(\mathbf{x^d}\right)$, e.g., from an orientation image, in which the image's pixels coincide with the Fourier grid), allow us to obtain an initial guess for the polarization field in direct space $\varphi_{ij}^{o}\left(\mathbf{x^d}\right)$ (Eq. 49), which in turn can be Fourier-transformed to obtain $\hat{\varphi}_{ij}^{o}\left(\mathbf{\xi^d}\right)$. Furthermore, assuming that $\lambda_{ij}^{o}\left(\mathbf{x^d}\right) = \sigma_{ij}^{o}\left(\mathbf{x^d}\right)$ is the initial guess for a field of Lagrange multipliers associated with the compatibility constraints, the iterative procedure based on augmented Lagrangians proposed by Michel et al. (2000) reads as follows. With the polarization field after iteration $n$ being known, the $n + 1$-th iteration starts by computing the new guess for the kinematically admissible strain-rate deviation field:

$$\hat{\dot{d}}_{ij}^{n+1}\left(\mathbf{\xi^d}\right) = -\hat{\Gamma}_{ijkl}^{sym}\left(\xi^d\right) \ \hat{\phi}_{kl}^{n}\left(\xi^d\right), \quad \forall\xi^d \neq \mathbf{0}; \quad \text{and} \quad \hat{\dot{d}}_{ij}^{n+1}\left(\mathbf{0}\right) = 0 \qquad (58)$$

where $\hat{\Gamma}_{ijkl}^{sym}$ is the Green operator, appropriately symmetrized. The corresponding field in real space is thus obtained by application of the inverse FFT, i.e.:

$$\tilde{d}_{ij}^{n+1}\left(\mathbf{x^d}\right) = \text{fft}^{-1}\left\{\hat{\dot{d}}_{ij}^{n+1}\left(\mathbf{\xi^d}\right)\right\} \qquad (59)$$

and the new guess for the deviatoric stress field is calculated from (omitting subindices):

$$\sigma^{\prime n+1}\left(\mathbf{x^d}\right) + \ \mathbf{L}^{o} : \dot{\gamma}_{o}\sum_{k}\mathbf{m}^{k}\left(\mathbf{x^d}\right) \ \left(\frac{\left|\mathbf{m}^{k}(\mathbf{x^d}):\sigma^{\prime n+1}(\mathbf{x^d})\right|}{\tau^{k}(\mathbf{x^d})}\right)^{n}$$
$$\times \text{sgn}\left(\mathbf{m}^{k}\left(\mathbf{x^d}\right) : \sigma^{\prime n+1}\left(\mathbf{x^d}\right)\right) = \lambda^{n}\left(\mathbf{x^d}\right) + \mathbf{L}^{o} : \left(\dot{\mathbf{E}} + \tilde{\mathbf{d}}^{n+1}\left(\mathbf{x^d}\right)\right) \qquad (60)$$

The iteration is completed with the calculation of the new guess of the Lagrange multiplier field:

$$\lambda^{n+1}\left(\mathbf{x^d}\right) = \lambda^{n}\left(\mathbf{x^d}\right) + \mathbf{L}^{o} : \left(\tilde{\dot{\varepsilon}}^{n+1}\left(\mathbf{x^d}\right) - \tilde{\mathbf{d}}^{n+1}\left(\mathbf{x^d}\right)\right) \qquad (61)$$

Equations (61) and (62) guarantee the convergence of $\dot{\varepsilon}\left(\mathbf{x^d}\right)$ (i.e., the strain-rate field related with the stress through the constitutive equation) toward $\mathbf{d}(\mathbf{x^d})$ (i.e., the kinematically admissible strain-rate field) to fulfill compatibility and the Lagrange multiplier field $\lambda(\mathbf{x^d})$ toward the stress field $\sigma^{\prime}(\mathbf{x^d})$ to fulfill equilibrium.

Upon convergence, the microstructure can be updated using an explicit scheme (see Lebensohn et al. 2008 for details).

## 3   Results

### 3.1   Validation of the Full-Field Formulation Using an Analytical Result

Let us consider a model polycrystal consisting of columnar orthorhombic grains with symmetry axes aligned with the $x_3$ axis, such that, when loaded in antiplane mode with shearing direction along $x_3$, the only two slip systems that can be activated in the grains are those defined by the following Schmid tensors:

$$\mathbf{m}^s = \left(\mathbf{e}_1 \otimes \mathbf{e}_3 + \mathbf{e}_3 \otimes \mathbf{e}_1\right)/2 \;\; , \quad \mathbf{m}^h = \left(\mathbf{e}_2 \otimes \mathbf{e}_3 + \mathbf{e}_3 \otimes \mathbf{e}_2\right)/2 \tag{62}$$

where $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is an orthonormal basis of crystallographic axes and "s" and "h" stand for soft and hard slip systems, respectively. If we further consider that $\mathbf{e}_3$ lies parallel to $x_3$, and the material is incompressible, the problem becomes two-dimensional (2-D). The local stress and strain rate are characterized by 2-D vectors with components $\sigma_{13}$ and $\sigma_{23}$ and $\mathbf{e}_3$ and $\dot{\varepsilon}_{23}$ (denoted hereafter $\sigma_1$ and $\sigma_2$ and $\dot{\varepsilon}_1$ and $\dot{\varepsilon}_2$, respectively) and the viscous stiffness tensor $\mathbf{L} = 2\mu$, by a 2-D symmetric second-order tensor with diagonal components $2\mu_{1313}$ and $2\mu_{2323}$ and off-diagonal components $2\mu_{1323}$ (denoted $2\mu_{11}$, $2\mu_{22}$, and $2\mu_{12}$, respectively). In addition, let us assume that the constituent grains exhibit a linear response:

$$\dot{\boldsymbol{\varepsilon}}\left(\mathbf{x}\right) = \mathbf{L}^{-1} : \boldsymbol{\sigma}\left(\mathbf{x}\right) = \left(\frac{1}{\tau_o^s}\mathbf{m}^s \otimes \mathbf{m}^s + \frac{1}{\tau_o^h}\mathbf{m}^h \otimes \mathbf{m}^h\right) : \boldsymbol{\sigma}\left(\mathbf{x}\right) \tag{63}$$

with $\tau_o^s$ and $\tau_o^h$ being the viscosities of the soft and hard slip systems $\left(\tau_o^s < \tau_o^h\right)$. It can be shown that the behavior of such polycrystal is characterized by an effective 2-D viscous stiffness tensor $\overline{\mathbf{L}} = 2\overline{\mu}$ such that $\dot{\mathbf{E}} = \overline{\mathbf{L}}^{-1} : \boldsymbol{\Sigma}$ (where $\boldsymbol{\Sigma}$ and $\dot{\mathbf{E}}$ are the 2-D effective stress and strain rate, respectively), such that (Dykhne 1970):

$$\det\left(\overline{\boldsymbol{\mu}}\right) = \overline{\mu}_{11} \times \overline{\mu}_{22} - \overline{\mu}_{12}^2 = \tau_o^s \times \tau_o^h \tag{64}$$

In the particular case of an isotropic 2-D polycrystal, $\overline{\mu}_{11} = \overline{\mu}_{22} \left(= \overline{\mu}\right)$ and $\overline{\mu}_{12} = 0$, so that the effective shear modulus becomes:

$$\overline{\mu} = \sqrt{\tau_o^s \times \tau_o^h} \tag{65}$$
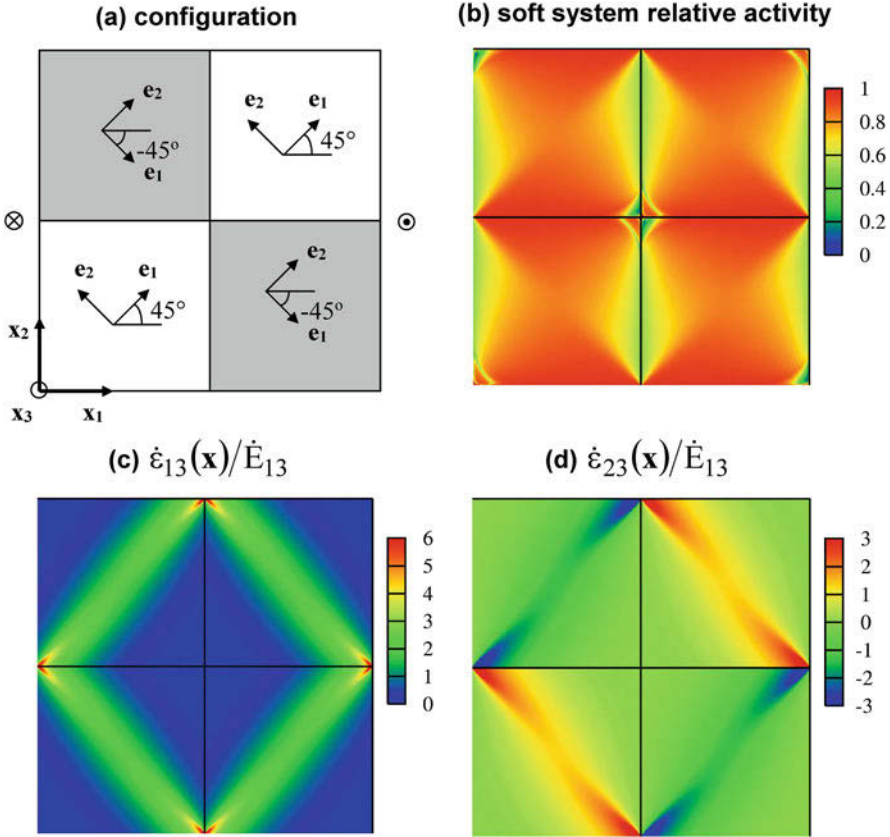
**Fig. 3** (**a**) Two-dimensional two-phase isotropic unit cell undergoing antiplane deformation. VPFFT results of (**b**) relative activity field of the soft slip system. (**c, d**) Shear components 13 and 23 of the strain rate

Note that the above result is independent of the 2-D microstructure as far as it remains isotropic. This analytical result can be used for validating the VPFFT formulation. Let us consider the periodic 2-D two-phase composite shown in Fig. 3a (Lebensohn et al. 2005), whose unit cell consists of four square grains, with the crystallographic orientations of the two pairs of opposite grains (i.e., each pair shearing only the central vertex) being characterized by angles $+45$ and $-45°$, respectively (note that the orientation of each 2-D crystal is fully characterized by the angle between the crystal direction $\mathbf{e}_1$ and the sample direction $x_1$).

The antiplane deformation of this unit cell for an applied strain rate of the form $\dot{\mathbf{E}} = (\dot{E}_{13}, 0)$ was solved numerically using different discretizations: 64, 128, 256, and 512 Fourier points along each direction (i.e., 1024, 4096, 16384, 65536 Fourier points per grain), for a contrast of $\tau_o^h/\tau_o^s = 25$, which gives theoretical polycrystal viscosity of $\overline{\mu}/\tau_o^s = 5$. Figure 4 shows the relative deviations of the

**Fig. 4** Relative deviations from the theoretical (th) value $(\overline{\mu}/\tau_o^s = 5)$ of the polycrystal's viscosity predicted with the present formulation (FFT) for different grid refinements, in the case 2-D antiplane deformation of the isotropic unit cell of Fig. 3a, consisting of grains having linear behavior (Eq. 63) with $\tau_o^h/\tau_o^s = 25$
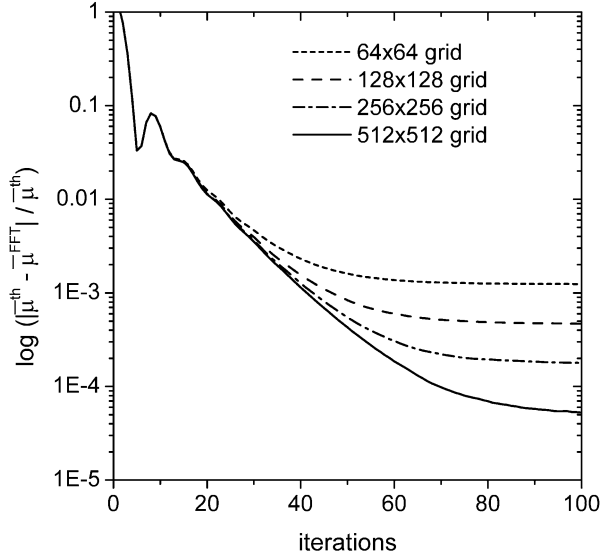


polycrystal viscosities calculated with the VPFFT model from the theoretical value, as the number of iterations of the VPFFT method increases. It is seen that (a) the convergence of $\overline{\mu}^{FFT}$ toward its theoretical value is rather good, although it saturates at different levels, depending on the number of discretization points used, and (b) the precision of the FFT solution can be increased by refining appropriately the Fourier grid. This is due to the fact that a more refined grid provides a higher spatial resolution to represent the strong gradients and jumps of the local fields, localized at grain boundaries (see Fig. 3c, d).

## 3.2    Validation of Mean-Field Formulations Using Full-Field Computations

In this section, we analyze the orientation dependence of the lattice rotation rate distributions given by Eqs. (45), (46), and (47) for the cases of axisymmetric deformation of fcc and bcc polycrystals (Lebensohn et al. 2016).

First, the statistical parameters describing the lattice rotation rate distributions predicted with VPSC are validated by comparison with corresponding full-field predictions obtained with the VPFFT model for the case of asymmetric tension of an fcc polycrystal with random texture and equiaxed grains.

Figure 5a shows VPSC predictions of the per-grain rotation rate averages and standard deviations normalized by $\overline{\dot{\omega}} = \langle \dot{\omega}^{(r)} \rangle$, corresponding to an fcc polycrystal made of 400 randomly oriented spherical grains deforming by {111}<110> slip and stress exponent $n = 10$, deformed axisymmetrically in tension at an applied strain rate of 1 s$^{-1}$. Figure 5b compares the same statistical parameters obtained
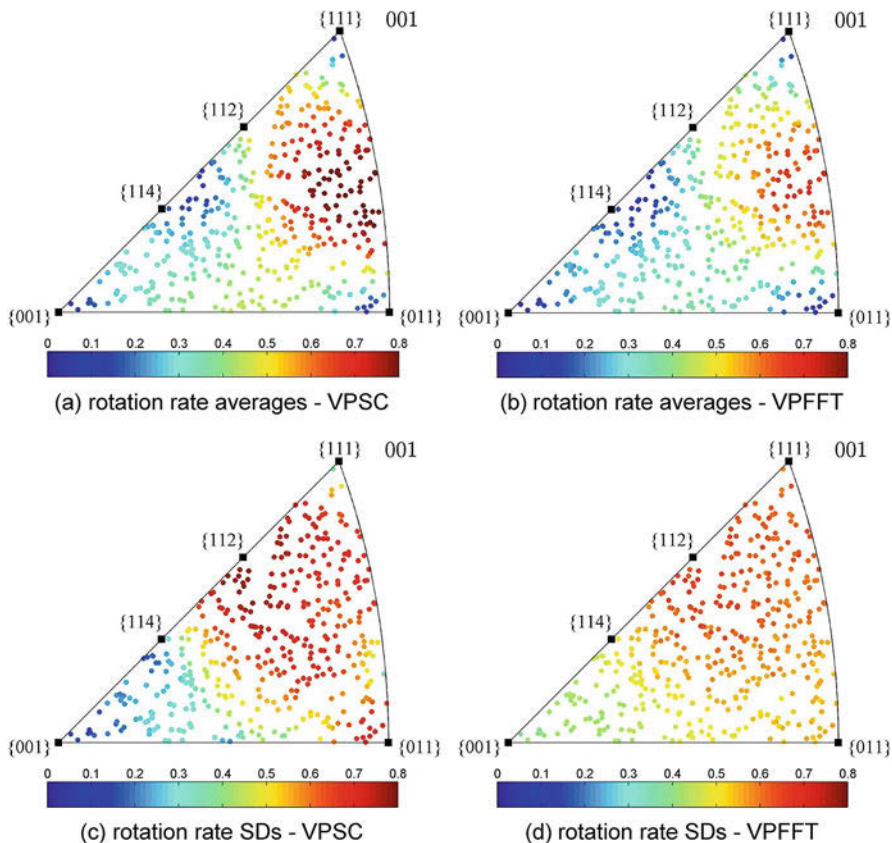
**Fig. 5** VPSC (left) and VPFFT (right) predictions of per-grain (**a**, **b**) rotation rate averages and (**c**, **d**) geometric mean standard deviations, for the case of tensile deformation of an fcc polycrystal made of 400 randomly oriented spherical/equiaxed grains deforming by {111}<110>

by sampling and ensemble averaging VPFFT predictions of the rotation rate fields, calculated for 50 different periodic Voronoi unit cell realizations, made of the same 400 orientations randomly assigned to the Voronoi grains, with identical constitutive behavior and applied boundary conditions.

The comparison between the VPSC predictions and the corresponding full-field statistics shows reasonable agreement. Both models predict very similar orientation dependence of the average rotation rates and, most importantly, of the average dispersions, with maximum near <112>, which incidentally coincides with the predominant recrystallization texture component of an fcc material after being heavily deformed in tension. The minor differences between VPSC and VPFFT

predictions can be ascribed to two reasons. (1) The unavoidable adoption of a linearization scheme in VPSC, affine in this case. As discussed above, this implies an approximation of the actual nonlinear response of the material, which the VPFFT solutions strictly account for. In the present case of relatively mild single crystal anisotropy and nonlinearity, the results of the first-order affine linearization match reasonable well the reference full-field solutions. Materials with higher local mechanical contrast would require the use of second-order linearization, for more accurate results. (2) The imperfect statistical representation of the distributions obtained by ensemble averaging of VPFFT solutions.

## 4  Summary and Further Reading

In this chapter, we have reviewed two crystal plasticity methodologies for the prediction of microstructure-property relations in polycrystalline aggregates. The mean-field VPSC and the full-field VPFFT formulations rely on Green's function-based solution of the PDEs governing the micromechanical response of polycrystals – a class of heterogeneous materials, in which the heterogeneity is related to the inherent directional properties of single crystal grains with respect to mechanical solicitation along different crystal directions and the different crystallographic orientations of each of these grains in the aggregate. The non-space-resolved VPSC formulation is based on a certain linearization assumption for the behavior of the single crystal grains, along with the integration – utilizing Fourier transforms – in an ellipsoidal domain of the Green's function for an infinite medium with homogenous properties and an iterative procedure to adjust the properties of this HEM. The space-resolved VPFFT method is based on the evaluation of convolution integrals over the entire periodic unit cell, between the Green's function for a periodic medium and a polarization term containing all the information on the heterogeneity and nonlinearity of the crystalline material's behavior. The numerical evaluation of such convolutions is performed in Fourier space and anti-transformed to Cartesian space by means of the efficient FFT algorithm. Both Green's function-based crystal plasticity formulations were presented using a unified notation, pointing out their similarities and differences and cross-validating their predictions.

The VPSC formulation has been extensively applied to predict microstructure-property relationships, including texture-induced anisotropy and microstructure evolution, both in stand-alone, single material point calculations and as part of multiscale frameworks, coupling the homogenization-based model with FE codes. Presenting these applications is beyond the scope of this chapter, but a list of stand-alone applications of VPSC to metals, minerals, and polymers can be found in Lebensohn et al. (2007), and the different multiscale strategies are discussed in, e.g., Tomé et al. (2001), Plunkett et al. (2006), Nixon et al. (2010), Segurado et al. (2012), and Barton et al. (2015).

Beyond the rigid-viscoplastic regime in which this chapter has been focused, the FFT-based formulation has been implemented for other deformation regimes, including polycrystal elasticity (Brenner et al. 2009), dilatational viscoplasticity (Lebensohn et al. 2013), and elasto-viscoplasticity (EVPFFT) (Lebensohn et al. 2012; Grennerat et al. 2012, for small strains; Eisenlohr et al. 2013, for finite strains). In particular, the standard EVPFFT formulation based on local plasticity was recently extended to consider strain-gradient and high-order plasticity theories, including field dislocation mechanics (FDM), by Berbenni et al. (2014), Brenner et al. (2014), and Lebensohn and Needleman (2016), etc.

# References

Barbe F, Decker L, Jeulin D, Cailletaud G (2001) Intergranular and intragranular behavior of polycrystalline aggregates. Part 1: F.E. model. Int J Plast 17:513–536

Barton NR, Bernier JV, Lebensohn RA, Boyce DE (2015) The use of discrete harmonics in direct multi-scale embedding of polycrystal plasticity. Comput Methods Appl Mech Eng 283: 224–242

Becker R (1991) Analysis of texture evolution in channel die compression. 1. Effects of grain interaction. Acta Metall Mater 39:1211–1230

Berbenni S, Taupin V, Djaka KS, Fressengeas C (2014) A numerical spectral approach for solving elasto-static field dislocation and g-disclination mechanics. Int J Solids Struct 51:4157–4175

Berveiller M, Fassi-Fehri O, Hihi A (1987) The problem of 2 plastic and heterogeneous inclusions in an anisotropic medium. Int J Eng Sci 25:691–709

Beyerlein IJ, Tomé CN (2008) A dislocation-based constitutive law for pure Zr including temperature effects. Int J Plast 24:867–895

Bobeth M, Diener G (1987) Static elastic and thermoelastic field fluctuations in multiphase composites. J Mech Phys Solids 35:137–149

Brenner R, Lebensohn RA, Castelnau O (2009) Elastic anisotropy and yield surface estimates of polycrystals. Int J Solids Struct 46:3018–3026

Brenner R, Beaudoin AJ, Suquet P, Acharya A (2014) Numerical implementation of static Field Dislocation Mechanics theory for periodic media. Phil Mag A 94:1764–1787

Brisard S, Dormieux L (2010) FFT-based methods for the mechanics of composites: a general variational framework. Comput Mater Sci 49:663–671

Cheng J, Ghosh S (2015) A crystal plasticity FE model for deformation with twin nucleation in magnesium alloys. Int J Plast 67:148–170

Cruzado A, Gan B, Jimenez M, Barba D, Ostolaza K, Linaza A, Molina-Aldareguia JM, LLorca J, Segurado J (2015) Multiscale modeling of the mechanical behavior of IN718 superalloy based on micropillar compression and computational homogenization. Acta Mater 98: 242–253

Delaire F, Raphanel JL, Rey C (2001) Plastic heterogeneities of a copper multicrystal deformed in uniaxial tension: experimental study and finite element simulations. Acta Mater 48:1075

Delannay L, Jacques PJ, Kalidindi SR (2006) Finite element modeling of crystal plasticity with grains shaped as truncated octahedrons. Int J Plast 22:1879–1898

Dykhne AM (1970) Conductivity of a two-dimensional two-phase system. Dokl Akad Nauk SSSR 59:110–115

Eisenlohr P, Diehl M, Lebensohn RA, Roters F (2013) A spectral method solution to crystal elasto-viscoplasticity at finite strains. Int J Plast 46:37–53

Eyre DJ, Milton GW (1999) A fast numerical scheme for computing the response of composites using grid refinement. Eur Phys J Appl Phys 6:41–47

Grennerat F, Montagnat M, Castelnau O, Vacher P, Moulinec H, Suquet P, Duval P (2012) Experimental characterization of the intragranular strain field in columnar ice during transient creep. Acta Mater 60:3655–3666

Hershey AV (1954) The elasticity of an isotropic aggregate of anisotropic cubic crystals. J Appl Mech 21:236–240

Hill R (1965) Continuum micro-mechanics of elastoplastic polycrystals. J Mech Phys Solids 13:89

Hutchinson JW (1976) Bounds and self-consistent estimates for creep of polycrystalline materials. Proc R Soc Lond A 348:101–127

Idiart MI, Moulinec H, Ponte Castaneda P, Suquet P (2006) Macroscopic behavior and field fluctuations in viscoplastic composites: second-order estimates versus full-field simulations. J Mech Phys Solids 54:1029–1063

Knezevic M, Lebensohn RA, Cazacu O, Revil-Baudard B, Proust G, Vogel S, Nixon ME (2013) Modeling bending of alpha-titanium with embedded polycrystal plasticity in implicit finite elements. Mat Sci Eng A 564:116–126

Kreher W (1990) Residual-stresses and stored elastic energy of composites and polycrystals. J Mech Phys Solids 38:115–128

Laws N (1973) On the thermostatics of composite materials. J Mech Phys Solids 21:9–17

Lebensohn RA (2001) N-site modelling of a 3D viscoplastic polycrystal using Fast Fourier Transform. Acta Mater 49:2723–2737

Lebensohn RA, Canova GR (1997) A selfconsistent approach for modelling texture development of two-phase polycrystals: application to Titanium alloys. Acta Mater 45:3687–3694

Lebensohn RA, Needleman A (2016) Numerical implementation of non-local polycrystal plasticity using Fast Fourier Transforms. J Mech Phys Solids 97:333–351

Lebensohn RA, Tomé CN (1993) A selfconsistent approach for the simulation of plastic deformation and texture development of polycrystals: application to zirconium alloys. Acta Metall Mater 41:2611–2624

Lebensohn R, Uhlenhut H, Hartig C, Mecking H (1998) Mechanical behavior γ-TiAl-based polysinthetically twinned crystals: micromechanical modelling and experimental validation. Acta Mater 46:4701–4709

Lebensohn RA, Castelnau O, Brenner R, Gilormini P (2005) Study of the antiplane deformation of linear 2-D polycrystals with different microstructures. Int J Solids Struct 42:5441–5449

Lebensohn RA, Tomé CN, Ponte Castañeda P (2007) Self-consistent modeling of the mechanical behavior of viscoplastic polycrystals incorporating intragranular field fluctuations. Philos Mag 87:4287–4322

Lebensohn RA, Brenner R, Castelnau O, Rollett AD (2008) Orientation image-based micromechanical modelling of subgrain texture evolution in polycrystalline copper. Acta Mater 56:3914–3926

Lebensohn RA, Idiart MI, Ponte Castañeda P, Vincent PG (2011) Dilatational viscoplasticity of polycrystalline solids with intergranular cavities. Philos Mag 91:3038–2067

Lebensohn RA, Kanjarla AK, Eisenlohr P (2012) An elasto-viscoplastic formulation based on fast Fourier transforms for the prediction of micromechanical fields in polycrystalline materials. Int J Plast 32&33:59–69

Lebensohn RA, Escobedo JP, Cerreta EK, Dennis-Koller D, Bronkhorst CA, Bingert J (2013) Modelling void growth in polycrystalline materials. Acta Mater 61:6918–6932

Lebensohn RA, Zecevic M, Knezevic M, McCabe RJ (2016) Average intragranular misorientation trends in polycrystalline materials predicted by a viscoplastic self-consistent approach. Acta Mater 104:228–236

Liu Y, Ponte Castañeda P (2004) Second-order theory for the effective behavior and field fluctuations in viscoplastic polycrystals. J Mech Phys Solids 52:467–495

Masson R, Bornert M, Suquet P, Zaoui A (2000) Affine formulation for the prediction of the effective properties of nonlinear composites and polycrystals. J Mech Phys Solids 48:1203

Michel JC, Moulinec H, Suquet P (2000) A computational method based on augmented Lagrangians and fast Fourier transforms for composites with high contrast. Comput Model Eng Sci 1:79–88

Mika DP, Dawson PR (1998) Effects of grain interaction on deformation in polycrystals. Mater Sci Eng A 257:62–76

Molinari A, Canova GR, Ahzi S (1987) Self consistent approach of the large deformation polycrystal viscoplasticity. Acta Metall 35:2983–2994

Moulinec H, Suquet P (1994) A fast numerical method for computing the linear and nonlinear mechanical properties of composites. C R Acad Sci Paris II 318:1417–1423

Moulinec H, Suquet P (1998) Numerical method for computing the overall response of nonlinear composites with complex microstructure. Comput Methods Appl Mech Eng 157:69–94

Mura T (1987) Micromechanics of defects in solids. Martinus-Nijhoff Publishers, Dordrecht

Musienko A, Tatschl A, Schmidegg K, Kolednik O, Pippan R, Cailletaud G (2007) Three-dimensional finite element simulation of a polycrystalline copper specimen. Acta Mater 55:4121–4136

Nixon ME, Cazacu O, Lebensohn RA (2010) Anisotropic response of high-purity α-titanium. Experimental characterization and constitutive modelling. Int J Plast 26:516–532

Parton VZ, Buryachenko VA (1990) Stress fluctuations in elastic composites. Sov Phys Dokl 35(2):191–193

Plunkett B, Lebensohn RA, Cazacu O, Barlat F (2006) Evolving yield function of hexagonal materials taking into account texture development and anisotropic hardening. Acta Mater 54:4159–4169

Ponte Castañeda P (2002) Second-order homogenization estimates for nonlinear composites incorporating field fluctuations: I- theory. J Mech Phys Solids 50:737–757

Proust G, Tomé CN, Kaschner GC (2007) Modeling texture, twinning and hardening evolution during deformation of hexagonal materials. Acta Mater 55:2137–2148

Quey R, Driver JH, Dawson P (2015) Intra-grain orientation distributions in hot-deformed aluminium: orientation dependence and relation to deformation mechanisms. J Mech Phys Solids 84:506–527

Raabe D, Sachtleber M, Zhao Z, Roters F, Zaefferer S (2001) Micromechanical and macromechanical effects in grain scale polycrystal plasticity experimentation and simulation. Acta Mater 49:3433–3441

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58:1152–1211

Sachs G (1928) On the derivation of a condition of flowing. Z Verein Deut Ing 72:734–736

Segurado J, Lebensohn RA, LLorca J, Tomé CN (2012) Multiscale modeling of plasticity based on embedding the viscoplastic self-consistent formulation in implicit finite elements. Int J Plast 28:124–140

Taylor GI (1938) Plastic strain in metals. J Inst Met 62:307–324

Tomé CN, Maudlin PJ, Lebensohn RA, Kaschner GC (2001) Mechanical response of zirconium. Part I: derivation of a polycrystal constitutive law and finite element analysis. Acta Mater 49:3085–3096

Wen W, Borodachenkova M, Tomé CN, Vincze G, Rauch EF, Barlat F, Gracio JJ (2016) Mechanical behavior of Mg subjected to strain path changes: experiments and modeling. Int J Plast 73:171–183

Wenk HR, Canova GR, Brechet Y, Flandin L (1997) A deformation-based model for recrystallization of anisotropic materials. Acta Mater 45:3283

Willot F (2015) Fourier-based schemes for computing the mechanical response of composites with accurate local fields. CR Mécanique 34:232–245

Zecevic M, Pantleon W, Lebensohn RA, McCabe RJ, Knezevic M (2017) Predicting intragranular misorientation distributions in polycrystalline metals using the viscoplastic self-consistent formulation. Acta Mater 140:398–410

# Computationally Efficient Crystal Plasticity Simulations Using Spectral Databases

# 78

Surya R. Kalidindi, Akash Gupta, and Evdokia Popova

## Contents

### Abstract

Crystal plasticity models allow incorporation of physics of plastic deformation at length scales smaller than a single crystal or individual grains in a polycrystalline

S. R. Kalidindi (✉) · E. Popova
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: surya.kalidindi@me.gatech.edu; evdokia.popova@gmail.com

A. Gupta
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA
TCS Research, Tata Consultancy Services, Tata Research Development and Design Centre, Pune, India
UMI 2958, Georgia Tech-CNRS, Metz, France
e-mail: 0309.akash@gmail.com

sample. As such, they present an important avenue for improving the fidelity of multiscale modeling and simulation for a variety of phenomena involving plastic deformations in crystalline materials. However, crystal plasticity models are extremely computationally expensive, limiting their adoption by materials development community and manufacturing industries. In this chapter, a spectral database approach is presented to carry out the computations involved in the crystal plasticity framework in a highly efficient manner. In this approach, the important variables of interest from crystal plasticity computations can be stored in precomputed databases based on discrete Fourier transforms (DFTs). Although the database generation requires one-time high computational cost, it dramatically reduces the computational cost in all subsequent computations. Several case studies illustrating the benefits of this approach are presented in this chapter.

# 1      Introduction

Modern manufacturing industry relies heavily on metal forming processes, which form the backbone of today's manufacturing world. Globally, hundreds of million tons of metals are subjected to metal forming processes every year. In metal forming processes (examples include rolling, extrusion, forging, deep drawing, and die forming), the products are shaped and often strengthened by plastic deformation. Plasticity also influences material failure (e.g., necking and fracture, fatigue). Consequently, a deeper understanding of plastic deformation in metals and alloys is central to efforts aimed at improving the manufacturing processes for optimizing the cost and the performance of the finished product.

Plasticity describes the deformation of a (solid) material undergoing nonreversible changes of shape in response to applied forces (Lubliner 2008). Most commercially used metals and alloys exhibit polycrystalline microstructures (material structure at the mesoscale) that are composed of numerous grains (individual crystals). In such materials, plastic deformation occurs predominantly by dislocation glide on specific crystallographic planes (typically close-packed planes) in specific crystallographic directions (typically close-packed directions) (Callister 2007). Consequently, the details of the material microstructure (such as the distribution of grain orientations, also referred as the crystallographic texture) play an important role in plasticity (Barlat 2007).

Most simulations of metal forming employ phenomenological descriptions of the constitutive laws controlling the macroscale (effective) elastic-plastic response of polycrystalline materials. These models are based on discrete macroscopic experimental observations and the assumption of material homogeneity. Although such models require relatively short computation times due to which they are the preferred choice for industrial process simulations, they are often not able to capture accurately the evolving anisotropy and properties (i.e., mechanical performance) of the finished product. For example, when it comes to processes involving ultra-thin sheet metals (few grains in the thickness), phenomenological models often

fail to capture material response accurately as these sheet metals present a strong heterogeneity of deformation during processing (Adzima et al. 2017). In such situations, it is essential to employ more physics-based descriptions of the material constitutive laws. For crystalline solids, these can be expressed in the form of crystal plasticity theories (Adams et al. 2013). A salient feature of these models is that they aim to express the constitutive relations of the material starting from the description of the key physical mechanisms at the single crystal scale. These constitutive relations account for the relevant microstructural mechanisms at this scale, such as slip on specific crystallographic planes and directions, the lattice rotation, and the dislocation motion and multiplication in the description of the strain hardening. The behavior at the macroscale is then subsequently derived from that of the constituent single crystals via suitable scale-transition schemes. Such micromechanical modeling approaches have demonstrated significant improvements in the predictions of the concomitant anisotropic mechanical response in polycrystalline materials and the evolution of the underlying texture in finite plastic deformation (e.g., Asaro and Needleman 1985; Kalidindi et al. 1992).

The transition from the response of single crystals to the overall (effective) response of the polycrystalline aggregate is usually accomplished using a suitable homogenization scheme. The commonly employed homogenization schemes can be classified based on the assumptions made with regards to the local interactions between grains, such as Taylor-type (also known as full constraints) (Taylor 1938), relaxed constraints (Kocks and Mecking 2003), LAMEL (Van Houtte et al. 2005), self-consistent (Molinari et al. 1987; Lebensohn et al. 2004; Lebensohn et al. 2007), and crystal plasticity finite element (Needleman et al. 1985; Kalidindi et al. 1992; Kalidindi and Anand 1994; Bachu and Kalidindi 1998; Kalidindi and Schoenfeld 2000; Raabe et al. 2005) models. The simplest and the most widely used approach is the Taylor-type model. In this method, the applied velocity gradient tensor at the microscale is assumed to be the same as the one applied at the macroscale (on the polycrystal). The macroscopic stress is obtained as a volume average over the entire polycrystalline aggregate. The Taylor-type model usually provides good predictions of the overall anisotropic stress-strain response and the averaged texture evolution for single-phase, high stacking faulty energy, cubic metals (Bronkhorst et al. 1992). However, it usually lacks good predictions at the scale of individual crystals and it fails to capture the development of heterogeneities within the grains (Kalidindi et al. 2004; Van Houtte et al. 2005). Crystal plasticity finite element method (called CPFEM) is a powerful modeling tool for a wide range of mechanical problems that take into account the local interactions between all grains in the sample. This approach uses the finite element (FE) method to find the response of the polycrystal by placing a finite element mesh over the grains such that each element represents one grain or a part of the grain (Roters et al. 2010).

The use of crystal plasticity models is extremely computationally expensive and has not been adopted broadly by the advanced materials development community. Even when the constitutive equations are developed and implemented efficiently, the time integration of the highly nonlinear single crystal response demands high computational resources. Consequently, in recent years, much attention has been devoted

to the development of novel techniques that reduce the computational cost involved, without significant loss of accuracy. The explored strategies have included both efficient numerical schemes and high-performance computing hardware (Knezevic and Kalidindi 2017). In the field of crystal plasticity modeling, these have included the Bunge–Esling approach (Bunge and Esling 1984; Kalidindi and Duvvuru 2005), Fourier transform (FFT)-based formulations (Lebensohn 2001; Lebensohn et al. 2008), Newton–Krylov method in place of the Newton–Raphson (NR) method for solving sets of highly nonlinear crystal plasticity equations (Chockalingam et al. 2013), and the database approaches which rely primarily on databases of precomputed solutions either in the form of Fourier (spectral) coefficients of the generalized spherical harmonics (GSH) (Kalidindi et al. 2006) or discrete Fourier transforms (DFTs) (Knezevic et al. 2008a; Knezevic and Savage 2014; Alharbi and Kalidindi 2015; Zecevic et al. 2015a). Among these, the database approaches have demonstrated the most dramatic reduction in the computational cost without any significant loss of accuracy. This is mainly because the database approaches retrieve the solutions to the single crystal constitutive equations from the precomputed databases instead of computing them directly (e.g., using the iterative NR method).

In this chapter, the DFTs-based database approach to speeding up the conventional crystal plasticity computations is presented and demonstrated using several exemplar case studies. More specifically, a database is designed and utilized to efficiently reproduce the solutions for the crystal plasticity constitutive descriptions. The versatility of this approach is demonstrated through a broad range of applications.

## 2    Crystal Plasticity Framework

In this section, the main details of rigid-viscoplastic crystal plasticity framework (Asaro and Needleman 1985) used in the Chapter are presented. Plastic deformation is assumed to be only due to the slip on the crystallographic slip systems. The velocity gradient, $\mathbf{L}$, at the single crystal scale can be expressed as

$$\mathbf{L} = \mathbf{W}^* + \mathbf{L}^P, \quad \mathbf{L}^P = \sum_\alpha \dot{\gamma}^\alpha \mathbf{m}^\alpha \otimes \mathbf{n}^\alpha \tag{1}$$

where $\mathbf{W}^*$ is the lattice spin tensor, $\mathbf{L}^P$ is the plastic velocity gradient tensor, $\dot{\gamma}^\alpha$ is the shearing rate on the slip system $\alpha$, and $\mathbf{m}^\alpha$ and $\mathbf{n}^\alpha$ denote the slip direction and the slip plane normal of the slip system $\alpha$, respectively, in the current configuration. The $\mathbf{m}^\alpha$ and $\mathbf{n}^\alpha$ are related to their counterparts $\mathbf{m}_0^\alpha$ and $\mathbf{n}_0^\alpha$, expressed in the isoclinic relaxed configuration by the lattice rotation tensor $\mathbf{R}^*$ as, $\mathbf{m}^\alpha = \mathbf{R}^* \cdot \mathbf{m}_0^\alpha$ and $\mathbf{n}^\alpha = \mathbf{R}^* \cdot \mathbf{n}_0^\alpha$ (Gupta et al. 2018). In the rate dependent formulation (Pan and Rice 1983), the shearing rate on each slip system depends on the resolved shear stress, $\tau^\alpha$, and the slip resistance, $s^\alpha$, of that slip system $\alpha$. It can be expressed in a power-law relationship as (Kalidindi et al. 1992):

$$\dot{\gamma}^\alpha = \dot{\gamma}_0 \left| \frac{\tau^\alpha}{s^\alpha} \right|^{\frac{1}{m}} \text{sgn}\left(\tau^\alpha\right), \quad \tau^\alpha = \boldsymbol{\sigma}' \cdot \mathbf{P}^\alpha \tag{2}$$

where $\dot{\gamma}_0$ is the reference value of the shearing rate, $m$ is the strain rate sensitivity parameter, $\boldsymbol{\sigma}'$ is the deviatoric part of the Cauchy stress tensor, and $\mathbf{P}^\alpha$ is the symmetric part of the Schmid tensor $\mathbf{m}^\alpha \otimes \mathbf{n}^\alpha$ given by

$$\mathbf{P}^\alpha = 0.5 \left(\mathbf{m}^\alpha \otimes \mathbf{n}^\alpha + \mathbf{n}^\alpha \otimes \mathbf{m}^\alpha\right) \tag{3}$$

Using Eqs. (1) and (3) stretching tensor $\mathbf{D}$ (the symmetric part of $\mathbf{L}$) is expressed as

$$\mathbf{D} = \sum_\alpha \dot{\gamma}^\alpha \mathbf{P}^\alpha \tag{4}$$

The evolution of the slip resistance can be described by a saturation-type law as follows:

$$\dot{s}^\alpha = h_0 \left(1 - \frac{s^\alpha}{s^s}\right)^a \sum_\beta |\dot{\gamma}^\beta| \tag{5}$$

where $h_0$, $s^s$, and $a$ denote the slip hardening parameters. Finally, the lattice spin tensor $\mathbf{W}^*$ (and the related lattice rotation tensor, $\mathbf{R}^*$ ) in the crystalline region is given by:

$$\mathbf{W}^* = \dot{\mathbf{R}}^* \mathbf{R}^{*\mathrm{T}} = \mathbf{W} - \mathbf{W}^\mathrm{p}, \qquad \mathbf{W}^\mathrm{p} = \sum_\alpha 0.5\dot{\gamma}^\alpha \left(\mathbf{m}^\alpha \otimes \mathbf{n}^\alpha - \mathbf{n}^\alpha \otimes \mathbf{m}^\alpha\right) \tag{6}$$

where $\mathbf{W}$ is the applied spin tensor, and $\mathbf{W}^\mathrm{p}$ is the plastic spin tensor.

## 3      Spectral Databases for Crystal Plasticity Computations in Cubic Metals

The crystal plasticity framework described in the previous section demands significant computational resources. The numerically stiff behavior of the crystal plasticity constitutive equations is a direct consequence of the fact that most metals have a very weak dependence on strain rate at room temperature, demanding the use of a small value for the strain rate sensitivity parameter in the flow rule (see Eq. (2)). Consequently, one typically needs several iterations to arrive at an acceptable solution. Additionally, the constitutive equations need to be solved several times in most crystal plasticity simulations. For example, the implementation of the crystal plasticity equations in a finite element tool requires solving the set of stiff constitutive equations for every crystal orientation at every integration point at every trial strain increment in the simulations. Consequently, the use of crystal

plasticity models for simulating practical engineering problems requires extremely high computational effort.

A spectral database approach (Knezevic et al. 2008a; Knezevic and Savage 2014; Alharbi and Kalidindi 2015; Zecevic et al. 2015a) has been developed for speeding up the crystal plasticity constitutive calculations. The main idea of this new approach is based on building a database of Fourier coefficients (based on DFTs) that efficiently capture the main solutions of the crystal plasticity constitutive formulations. As mentioned earlier, we intend to use such a database in performing all calculations in a multiscale crystal plasticity simulation, avoiding the need to solve directly the crystal plasticity equations.

The new database approach seeks compact representations for the following three functions corresponding to the main variables in the crystal plasticity solutions: (i) the deviatoric stress tensor $\boldsymbol{\sigma}'(g, \mathbf{L})$, (ii) the plastic spin tensor $\mathbf{W}^p(g, \mathbf{L})$, and (iii) the total shear rate $\sum_\alpha |\dot{\gamma}^\alpha| (g, \mathbf{L})$, where $g$ denotes the crystal lattice orientation represented using Bunge–Euler angles, and $\mathbf{L}$ is the traceless applied velocity gradient tensor. However, the space of $\mathbf{L}$ (i.e., the set of all possible values assigned to $\mathbf{L}$) is unwieldy and needs to be suitably parametrized. We seek parametrization of the velocity gradient as (Van Houtte 1994):

$$\mathbf{L} = \mathbf{D} + \mathbf{W}, \quad \mathbf{D} = \dot{\varepsilon}\mathbf{D}_0, \mathbf{D}_0 = \sum_{j=1}^{3} D_j \boldsymbol{e}_j^p \otimes \boldsymbol{e}_j^p, \quad \dot{\varepsilon} = \|\mathbf{D}\| \tag{7}$$

$$D_1 = \sqrt{\frac{2}{3}} \cos\left(\theta - \frac{\pi}{3}\right), \quad D_2 = \sqrt{\frac{2}{3}} \cos\left(\theta + \frac{\pi}{3}\right), \quad D_3 = -\sqrt{\frac{2}{3}} \cos(\theta) \tag{8}$$

where $\left\{\boldsymbol{e}_j^p, \ j = 1, 2, 3\right\}$ denotes the principal frame of the stretching tensor $\mathbf{D}$, and $\{D_j, \ j = 1, 2, 3\}$ are the components of a diagonal and traceless $\mathbf{D}_0$. It should be noted that none of the functions of interest identified above are dependent on $\mathbf{W}$ (can be seen from the constitutive model described in Sect. 2). As a result of the above parametrization, the functions of interest can then be expressed in terms of $\left(\left\{\boldsymbol{e}_j^p\right\}, \theta, \dot{\varepsilon}\right)$ instead of $\mathbf{L}$. For example, the function of interest for the deviatoric stress tensor can be expressed as $\boldsymbol{\sigma}'\left(g, \left\{\boldsymbol{e}_j^p\right\}, \theta, \dot{\varepsilon}\right)$.

One approach to reducing the number of independent variables in the functions is to select a specific reference frame where the components of the tensors will be described. For example, we could decide that the function will be described in the principal frame of $\mathbf{D}$. Making this choice implies that we would have to use a suitable coordinate transformation law to recover the values of the function of interest in a different reference frame (such as the sample reference frame). Selecting the principal frame of $\mathbf{D}$ as the reference frame also implies that the crystal lattice orientation has to be defined with respect to this frame (note that $g$ denotes the crystal lattice orientation with respect to the sample reference frame). Let $g^p$ denote

the crystal orientation with respect to the principal frame of $\mathbf{D}$ (defined by $\left\{e_j^p\right\}$). With the choices made above, the functions of interest can now be expressed over $(g^p, \theta, \dot{\varepsilon})$. As an example, the expected functional dependence of the components of the deviatoric stress tensor in the $\left\{e_j^p\right\}$ reference frame can be expressed as $\sigma\prime_{ij}(g^p, \theta, \dot{\varepsilon})$. The central significance of the rationale described above is that the number of independent input parameters for the functions of interest has been reduced from 11 to 5. More importantly, of the five independent input parameters, four are angles (three for $g^p$ and one for $\theta$). This is very significant because the angular variables have naturally periodic domains (often expressed as $[0, 2\pi)$) and are ideally suited for DFT representations. It is also pointed that although $\dot{\varepsilon}$ is not defined on a bounded space, it will not present a hurdle for our application, as we will be able to express the dependence of the functions of interest on this variable in explicit (i.e., analytical) forms.

The orientation variable $g^p$ is most conveniently described by a set of three Bunge–Euler angles denoted as $(\varphi_1, \Phi, \varphi_2)$. Since each of this variable is an angle, the natural domain for the orientation can be expressed in the most general case as $(\varphi_1 \epsilon [0, 2\pi), \Phi \epsilon [0, 2\pi), \varphi_2 \epsilon [0, 2\pi))$. Although the crystal and sample symmetries generally identify as significantly smaller space as the fundamental zone (i.e., the set of all distinct crystal orientations), it is most convenient to stay with the full orientation space identified above to take advantage of the fact that this space is naturally periodic and therefore once again ideally suited for DFT representations. For the variable $\theta$, we can identify periodic domain as $\theta \epsilon [0, 2\pi/3)$ (see Fig.1).

The DFT representations of the functions of interest mentioned above are expressed as:



**Fig. 1** Variation of components $D_1$, $D_2$ and $D_3$ as a function of $\theta$
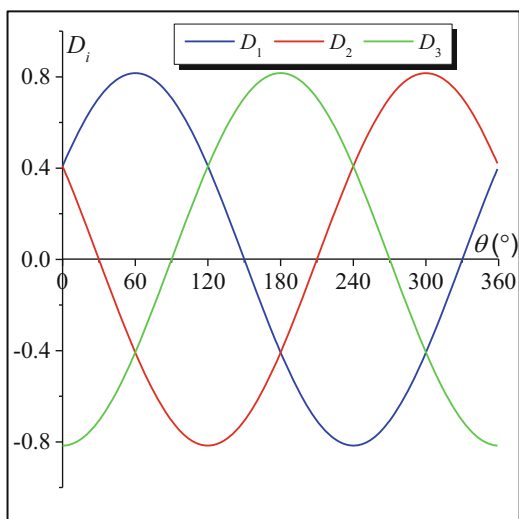
$$\mathbf{W}_{rutq}^p = \dot{\varepsilon}\, \frac{1}{N_{\varphi_1}\, N_\varphi N_{\varphi_2} N_\theta} \sum_{k=0}^{N_{\varphi_1}-1} \sum_{l=0}^{N_\varphi-1} \sum_{m=0}^{N_{\varphi_2}-1} \sum_{n=0}^{N_\theta-1} B_{klmn} e^{\frac{2\pi i k r}{N_{\varphi_1}}} e^{\frac{2\pi i l u}{N_\varphi}} e^{\frac{2\pi i m t}{N_{\varphi_2}}} e^{\frac{2\pi i n q}{N_\theta}}$$

(9)

$$\boldsymbol{\sigma}_{rutq}' = s\, \dot{\varepsilon}^m\, \frac{1}{N_{\varphi_1}\, N_\varphi N_{\varphi_2} N_\theta} \sum_{k=0}^{N_{\varphi_1}-1} \sum_{l=0}^{N_\varphi-1} \sum_{m=0}^{N_{\varphi_2}-1} \sum_{n=0}^{N_\theta-1}$$

$$C_{klmn} e^{\frac{2\pi i k r}{N_{\varphi_1}}} e^{\frac{2\pi i l u}{N_\varphi}} e^{\frac{2\pi i m t}{N_{\varphi_2}}} e^{\frac{2\pi i n q}{N_\theta}}$$

(10)

$$\left( \sum_\alpha |\dot{\gamma}^\alpha| \right)_{rutq} = \dot{\varepsilon}\, \frac{1}{N_{\varphi_1}\, N_\varphi N_{\varphi_2} N_\theta} \sum_{k=0}^{N_{\varphi_1}-1} \sum_{l=0}^{N_\varphi-1} \sum_{m=0}^{N_{\varphi_2}-1} \sum_{n=0}^{N_\theta-1}$$

$$G_{klmn} e^{\frac{2\pi i k r}{N_{\varphi_1}}} e^{\frac{2\pi i l u}{N_\varphi}} e^{\frac{2\pi i m t}{N_{\varphi_2}}} e^{\frac{2\pi i n q}{N_\theta}}$$

(11)

In Eqs. (9), (10), and (11), $r$, $u$, $t$, and $q$ enumerate the grid points in the orientation space (domain of $g^p$ or $(\varphi_1, \Phi, \varphi_2)$) and the $\theta$ space (describing the deformation mode), respectively, while $N_{\varphi_1}$, $N_\varphi$, $N_{\varphi_2}$, and $N_\theta$ represent the corresponding total number of grid points in the same spaces. In other words, the variables $r$, $u$, $t$, and $q$ take positive integer values. The sets of Fourier coefficients $B_{klmn}$, $C_{klmn}$, $G_{klmn}$ are generally referred to as the DFTs and constitute the spectral databases needed for facilitating the fast crystal plasticity solutions described here. Equations (9), (10), and (11) are given for three components ($W_{12}^p$, $W_{13}^p$, $W_{23}^p$) of plastic spin tensor, six components ($\sigma'_{11}$, $\sigma'_{22}$, $\sigma'_{33}$, $\sigma'_{12}$, $\sigma'_{13}$, $\sigma'_{23}$) of deviatoric stress tensor, and total shear rate, respectively. Note also that Eqs. (9), (10), and (11) have explicitly accounted for the dependence of the different functions on the variable $\dot{\varepsilon}$, which can be derived directly from the crystal plasticity theory presented earlier in Sect. 2. As a result, the DFT representations in these equations involve only a four-dimensional space. To incorporate hardening, slip resistance $s$ is included in Eq. (10) for calculation of the deviatoric stress, with the assumption $s = s^\alpha$ (i.e., the slip resistance is same for all the slip systems within a grain but varies from grain-to-grain). At the beginning of the time integration, the slip resistance is generally assigned the same initial value for all the grains. For subsequent time steps, the total shear rate calculated from Eq. (11) is used in Eq. (5) to update the slip resistance ($s^\alpha$) which determines the amount of strain hardening on slip system α for the next strain increment. An important consequence of this strategy is that the database of Fourier coefficients established in Eqs. (9), (10), and (11) is completely independent of the values of the hardening parameters. In fact, the Fourier coefficients are also independent of the exact form of the hardening equation as long as we make the assumption that all the slip systems exhibit the same slip resistance.

In order to build the spectral databases, we need to sample the functions of interest on a uniform grid in their domains. This sampling can be accomplished

by carrying out the crystal plasticity computations for the selected inputs using the model presented in Sect. 2. This does entail a significant computational cost. However, the main benefit of the approach presented here is that building the database constitutes a one-time cost, i.e., there is no need to rebuild this database because the database designed here is applicable to any fcc crystal subjected to any deformation path and exhibiting a broad range of hardening responses. In order to compute the DFTs, we need to sample the functions of interest on a uniform grid in 4-D (defined by $(\varphi_1, \Phi, \varphi_2, \theta)$). One has to decide on the resolution of the sampling grid. Obviously, finer grids will result in higher accuracy but can increase the (one-time) computational cost of generating the database. Furthermore, below a certain level of discretization, one would expect that there would be no new peaks in the frequency spectrum of the function of the interest, especially for the physics-based functions of interest identified above (i.e., these are expected to vary smoothly over their domain spaces). Multiple trials have shown that the crystal plasticity spectral databases produced with the one-degree grid discretization (Gupta et al. 2018) provide excellent results. Although there is an additional computation cost in calculating the database on the finer one-degree grid, it should be remembered that it is only a one-time expense. The periodic domain of independent variables on 4-D grid is chosen as $(\varphi_1 \epsilon [0, 2\pi), \Phi \epsilon [0, 2\pi), \varphi_2 \epsilon [0, 2\pi)), \theta \epsilon [0, 2\pi/3)$; thus, for one-degree discretization, it means that total simulations required to generate initial data for building database are 360 x 360 x 360 x 120 $\approx 6 \times 10^9$ simulations. This is a huge number of simulations, which will require significant computational resources and time even if simulations are run in parallel. We, therefore, explore ideas that can reduce the total number of required simulations.

The number of simulations required for generating DFT database can be reduced by taking advantage of the symmetry relations for cubic crystals (Adams et al. 2013) and the mirror symmetry evident within the periodic domain of $\theta$ (refer Fig.1). Simulations using Taylor-type model are done only for orientations in FZ3 (3 times fundamental zone of cubic crystals), and symmetry relations are used to fill up entire grid for generating DFTs. FZ3 is defined as (Adams et al. 2013)

$$FZ3 = \left\{ g = (\varphi_1, \Phi, \varphi_2) \left| \begin{array}{c} \left(0 \leq \varphi_1 < 2\pi, \ 0 \leq \Phi < \frac{\pi}{2}, \ 0 \leq \varphi_2 < \frac{\pi}{2}\right) \\ \left(0 \leq \varphi_1 < \pi, \ \Phi = \frac{\pi}{2}, \ 0 \leq \varphi_2 < \frac{\pi}{2}\right) \end{array} \right| \right\}$$

(12)

For the chosen one-degree grid, middle of the bin is chosen to be grid point. That is, grid points are 0.5, 1.5, 2.5, and so on; therefore, single point simulations using Taylor-type model were done to get initial data for grid points spaced in one degree increments over the space $(\varphi_1 \epsilon [0.5, 359.5], \Phi \epsilon [0.5, 89.5], \varphi_2 \epsilon [0.5, 89.5]), \theta \epsilon [0.5, 59.5]$. Therefore, total number of simulations required reduces to $360 \times 90 \times 90 \times 60 \approx 2 \times 10^8$, which is only 3% of the simulations required for the total periodic 4-D grid identified earlier.

The parameters for conventional crystal plasticity model for these simulations were chosen as: slip hardening parameter $(h_o) = 0$, initial slip resistance $(s_o) =$

100 MPa, reference value of the shearing rate $(\dot{\gamma}_o) = 0.001/s$, rate-sensitivity $(m) = 0.01$, strain rate $(\dot{\varepsilon}) = 0.001$. As noted earlier, the functions of interest were sampled for the input values selected on a uniform grid by performing these simulations. While generating the database hardening variable $s$ is kept constant and established Fourier coefficients database is independent of the values of the strain hardening parameters in Eq. (5).

Simulation results corresponding to each orientation in FZ3 will have to be assigned (copied) to fifteen other orientations which are equivalent to it before carrying out the DFTs. This is because there are sixteen equivalent orientations in the periodic grid; eight of these are obtained from crystal lattice symmetry operations and another eight directly from the definitions of the Bunge-Euler angles. The definitions of the Bunge-Euler angles require that locations $(\varphi_1, \Phi, \varphi_2)$ and $(\varphi_1 + \pi, 2\pi - \Phi, \varphi_2 + \pi)$ correspond to the exact same crystal lattice orientation for crystals of any symmetry. Furthermore, consideration of eight of the lattice symmetry operations associated with a cubic crystal lattice (e.g., face-centered cubic (fcc), body-centered cubic (bcc)) requires that locations corresponding to $(\varphi_1 + \pi, \pi - \Phi, 2\pi - \varphi_2)$, $(\varphi_1 + \pi, \pi - \Phi, \pi - \varphi_2)$, $(\varphi_1 + \pi, \pi - \Phi, \pi/2 - \varphi_2)$, $(\varphi_1 + \pi, \pi - \Phi, 3\pi/2 - \varphi_2)$, $(\varphi_1, \Phi, \varphi_2 + \pi/2)$, $(\varphi_1, \Phi, \varphi_2 + \pi)$ and $(\varphi_1, \Phi, \varphi_2 + 3\pi/2)$ also correspond to the exact same crystal lattice orientation defined by $(\varphi_1, \Phi, \varphi_2)$ (Adams et al. 2013). For deformation mode angle $\theta$, mirror symmetry of periodic domain is used to fill the remaining grid $(f(\theta) = f(120 - \theta))$. Thus, after considering above symmetry relations, we can fill up entire 4-D periodic grid of size 360 x 360 x 360 x 120 for all functions of interest. Since 4-D matrices for doing DFTs are large, it requires a large amount of computer memory (RAM) $\sim$ 128 GB to calculate DFTs and to perform further operations on them such as sorting. DFTs for each output variable (i.e., the functions of interest in crystal plasticity computations identified earlier) are converted from a 4-D matrix to a vector and stored in descending order of their magnitude with corresponding frequencies in a separate matrix. It should be noted that for each output variable, the size of the vector containing all DFTs is 5,598,720,000 $(= 360 \times 360 \times 360 \times 120)$ which requires $\sim$ 40 GB storage space. Therefore, these DFTs are sorted in descending order of magnitude, and only a reasonable number of the dominant DFTs are stored for further use. As an example, the reduced database for one output variable containing a vector of 50,000 DFTs requires $\sim$ 700 KB storage space. In producing such a compact set, we have also taken advantage of the fact that about half of the DFTs are complex conjugates of the other half for all real-valued functions. Considering the size of matrices and memory requirement, it is recommended to do all the steps together starting from populating the 4-D periodic grid from single point simulation data to storing sorted set of dominant DFTs. After obtaining the simulation data from conventional crystal plasticity model, it took a computational time of $\sim$70 h on a 2 GHz clock speed machine with a memory (RAM) of $\sim$ 128 GB to generate the complete DFT database for all output variables of interest (a total of ten). From the computed DFTs, the DFTs of interest defined in Eqs. (9), 10), and (11) are extracted.

Note that the DFTs were generated independently for all six components of the deviatoric stress, even though one of the components is redundant. For example, one of the components can be computed using the relation $\sigma'_{33} = -\left(\sigma'_{11} + \sigma'_{22}\right)$ to ensure that the trace of deviatoric stress tensor is zero. Our trials have shown that the strategy of building the database only for the five independent components and then computing the sixth component by setting the trace of the tensor to zero resulted in a larger error for $\sigma'_{33}$. This was a consequence of the addition of the truncation errors for $\sigma'_{11}$ and $\sigma'_{22}$, in computing $\sigma'_{33} = -\left(\sigma'_{11} + \sigma'_{22}\right)$. A better strategy is to establish an independent database for $\sigma'_{33}$ (Knezevic et al. 2009; Alharbi and Kalidindi 2015). For imposing the condition that the trace of the deviatoric stress tensor has to be zero, the hydrostatic part was subtracted from each of the normal stress components, after computing the full tensor using the spectral database. This approach produced much more accurate predictions of the entire deviatoric stress tensor, as will be evident from the results presented later.

For the estimation of the function values at points other than the grid points, there exist several methods that utilize the DFT representations of the function. All of these approaches implicitly assume that the original grid employed in computing the DFTs is sufficiently refined to capture all of the important frequencies embedded in the function of interest and perform some sort of an interpolation with global support. One approach is the spectral interpolation approach (Knezevic et al. 2009; Alharbi and Kalidindi 2015), where the DFTs are suitably padded to recover the function values on any finer grid on the same domain (Knezevic et al. 2008a). This methodology is known to be insensitive to the phenomenon of aliasing, which arises when trying to recover function values at locations other than those originally sampled. The DFT corresponding to the finer grid can then be used to compute efficiently an interpolated value of the function at any of the grid points on the finer grid. However, it was observed that spectral interpolation introduced certain rounding errors, which can be eliminated by the use of trigonometric interpolation, which is a continuous interpolation scheme (Zecevic et al. 2015a). Using the trigonometric interpolation scheme resulted in excellent accuracy of the crystal plasticity computations, as demonstrated in recent studies (Zecevic et al. 2015a; Gupta et al. 2018). The details of this interpolation scheme are presented next.

The main functions of crystal plasticity solutions are recovered as continuous functions of orientation and deformation mode from spectral DFT database with trigonometric interpolation using Inverse Discrete Fourier Transforms (IDFTs) equations in principal frame of $\mathbf{D}$, given by (Zecevic et al. 2015a):

$$
\mathbf{W}^p\left(g^p, \theta\right) = \dot{\varepsilon}\,\frac{1}{N_{\varphi_1} N_\varphi N_{\varphi_2} N_\theta}
$$

$$
\left[\sum_{k=-\frac{N_{\varphi_1}}{2}+1}^{\frac{N_{\varphi_1}}{2}} \sum_{l=-\frac{N_\varphi}{2}+1}^{\frac{N_\varphi}{2}} \sum_{m=-\frac{N_{\varphi_2}}{2}+1}^{\frac{N_{\varphi_2}}{2}} \sum_{n=-\frac{N_\theta}{2}+1}^{\frac{N_\theta}{2}} B_{klmn} e^{\frac{2\pi i k \varphi_1}{L_{\varphi_1}}} e^{\frac{2\pi i l \varphi}{L_\varphi}} e^{\frac{2\pi i m \varphi_2}{L_{\varphi_2}}} e^{\frac{2\pi i n \theta}{L_\theta}}\right]
$$

$$(13)$$

$$\sigma'\left(g^p, \theta\right) = s\dot{\varepsilon}^m \frac{1}{N_{\varphi_1} N_\varphi N_{\varphi_2} N_\theta}$$

$$\left[ \sum_{k=-\frac{N_{\varphi_1}}{2}+1}^{\frac{N_{\varphi_1}}{2}} \sum_{l=-\frac{N_\varphi}{2}+1}^{\frac{N_\varphi}{2}} \sum_{m=-\frac{N_{\varphi_2}}{2}+1}^{\frac{N_{\varphi_2}}{2}} \sum_{n=-\frac{N_\theta}{2}+1}^{\frac{N_\theta}{2}} C_{klmn} e^{\frac{2\pi i k \varphi_1}{L_{\varphi_1}}} e^{\frac{2\pi i l \varphi}{L_\varphi}} e^{\frac{2\pi i m \varphi_2}{L_{\varphi_2}}} e^{\frac{2\pi i n \theta}{L_\theta}} \right]$$

(14)

$$\sum_\alpha \mid \dot{\gamma}^\alpha \mid \left(g^p, \theta\right) = \dot{\varepsilon} \frac{1}{N_{\varphi_1} N_\varphi N_{\varphi_2} N_\theta}$$

$$\left[ \sum_{k=-\frac{N_{\varphi_1}}{2}+1}^{\frac{N_{\varphi_1}}{2}} \sum_{l=-\frac{N_\varphi}{2}+1}^{\frac{N_\varphi}{2}} \sum_{m=-\frac{N_{\varphi_2}}{2}+1}^{\frac{N_{\varphi_2}}{2}} \sum_{n=-\frac{N_\theta}{2}+1}^{\frac{N_\theta}{2}} G_{klmn} e^{\frac{2\pi i k \varphi_1}{L_{\varphi_1}}} e^{\frac{2\pi i l \varphi}{L_\varphi}} e^{\frac{2\pi i m \varphi_2}{L_{\varphi_2}}} e^{\frac{2\pi i n \theta}{L_\theta}} \right]$$

(15)

where $N_{\varphi_1}, N_\varphi, N_{\varphi_2}, N_\theta$ are the total number of grid points and $L_{\varphi_1}, L_\varphi, L_{\varphi_2}, L_\theta$ represent the domain of variables in the orientation space ($g^p$ or ($\varphi_1, \Phi, \varphi_2$)) and the deformation mode space ($\theta$). For one-degree grid, $N_{\varphi_1} = N_\varphi = N_{\varphi_2} = 360$, $N_\theta = 120$ and $L_{\varphi_1} = L_\varphi = L_{\varphi_2} = 360, L_\theta = 120$, since the periodic domain of variables considered are $\varphi_1 \epsilon [0, 360^o), \Phi \epsilon [0, 360^o), \varphi_2 \epsilon [0, 360^o), \theta \epsilon [0, 120^o)$. When retrieving Fourier coefficients ($B_{klmn}, C_{klmn}, G_{klmn}$) referred as DFTs for use in above equations to reconstruct values of above functions, we use only dominant DFTs and also take advantage of the fact that about half of the DFTs are complex conjugates of other half for all real valued functions. In the DFT-based spectral approach used to solve the crystal plasticity constitutive equations, only a small number of terms in the DFT (called dominant DFTs) are required to reconstruct directly the solutions for the main functions of the conventional crystal plasticity theory for any given crystal orientation and under any applied deformation mode. The use of small number of dominant DFTs speeds up crystal plasticity calculations by about two orders of magnitude without significantly compromising accuracy (Knezevic et al. 2009). Even with using one-degree grid discretization as in case study presented here (Gupta et al. 2018), the cost of recovering the solutions from database is roughly the same as the three-degree grid database (Knezevic et al. 2009), which is about two orders of magnitude faster than the conventional computations.

Once the spectral database is developed, it can be used for various applications as described in Sects. 4 and 6. Given a set of crystal orientations, an initial value of slip resistance per crystal orientation, hardening parameters, and an imposed value of the strain rate tensor, we can use Eqs. (13), (14), and (15) to compute required microscale quantities (i.e., grain scale stress, grain spin, and shearing rates on slip systems). A desired deformation process can be simulated by applying a suitable sequence of strain increment steps. The computed spin tensor for each crystal is used to update its lattice orientation. The updated set of crystal lattice orientations

are then used to define the starting conditions for the next strain increment. Total shear rate is used to update the slip resistance using Eq. (5) for the next strain increment. For the polycrystalline response (i.e., macroscale quantities such as the homogenized stress–strain response and texture), the Taylor model is used, in which the applied strain rate tensor at the grain level is assumed to be the same as the one applied at the polycrystal level. The macroscopic stress for the polycrystal is then obtained by volume averaging the stresses inside the polycrystal.

## 4   Application of Spectral Databases for Crystal Plasticity Computations of Cubic Metals

### 4.1   Taylor-Type Crystal Plasticity Simulations Using Spectral Database

In this section, case study of Taylor-type crystal plasticity simulations using spectral database built in Sect. 3 is presented. Uniaxial tension simulations were performed, and the hardening parameters were taken from Kalidindi et al. (1992) and correspond to oxygen-free high conductivity (OFHC) copper: $h_0 = 180$ MPa, $s_s = 148$ MPa, $a = 2.25$, $s_0 = 16$ MPa, and $m = 0.01$. The polycrystal was assumed to consist of 1000 FCC single crystals possessing a random initial texture. A comparative analysis of results predicted by the one-degree and three-degree grid databases is also performed.

The results of uniaxial tension simulations are presented in Fig. 2. The averaged polycrystal responses are plotted in this figure as macroscale deviatoric stress-strain curves. Deformed textures (pole figures) obtained using conventional crystal plasticity Taylor-type model (CP) and DFT database (SCP) show an excellent agreement. Comparison of deviatoric stress-strain curves of CP and SCP using one-degree grid and three-degree grid show that one-degree grid provides more accurate predictions. It should be noted that simulations are rigid-viscoplastic. For SCP model, 500 dominant DFTs were chosen for deviatoric stress components ($C_{klmn}$) and total shear rate ($G_{klmn}$) and 2000 dominant DFTs for plastic spin components ($B_{klmn}$) with a strain step of 0.06.

Computational cost of DFT database approach improved by two orders of magnitude compared to conventional crystal plasticity simulation without significantly compromising accuracy. It should be noted that we can further improve the accuracy of SCP model by using more number of dominant DFTs, but then computational cost will be higher. The number of dominant DFTs chosen here represents a compromise between computational time and accuracy.

### 4.2   Fast Computation of Yield Surfaces Using Spectral Databases

In this section, a new method for the fast computation of the yield surface in the five-dimensional deviatoric stress space is described using the DFT-based databases. Since the computational cost of delineation of the anisotropic yield surface in stress
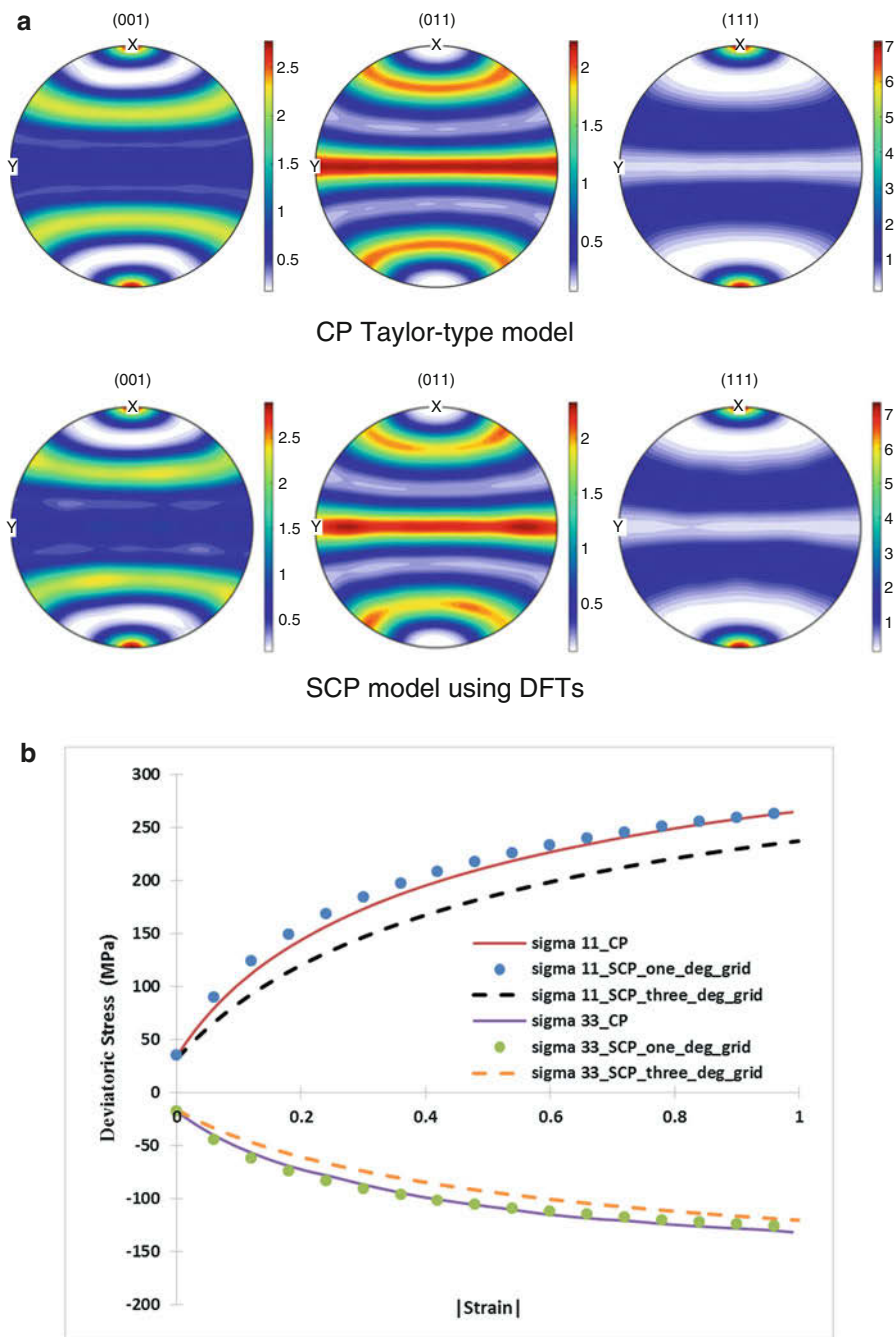
**Fig. 2** Comparison of predictions from SCP model using one-degree grid and three-degree grid database against corresponding predictions from conventional CP Taylor-type model for uniaxial tension simulation of OFHC Copper: (**a**) pole figures of deformed texture and (**b**) different components (sigma 11 and sigma 33) of deviatoric stress vs. strain curves

space using crystal plasticity is very high, the spectral database approach can be used to tackle this problem.

The orientation distribution function (ODF), denoted as $f(g)$, reflects the normalized probability density associated with occurrence of the crystallographic orientation, $g$, in the sample. ODF is formally defined as

$$f(g)dg = \frac{N_{g \pm dg/2}}{N}, \qquad \int_{FZ} f(g)dg = 1 \tag{16}$$

where $N$ is the total number of orientations measured in the sample, $N_{g \pm dg/2}$ is the number of orientations that lie within an invariant measure, $dg$, centered about the orientation, $g$, and $FZ$ denotes the fundamental zone of distinct Bunge-Euler angles. The invariant measure is then defined as:

$$dg = \sin \varphi \, d\varphi_1 d\varphi d\varphi_2 \tag{17}$$

In this work, it is convenient to establish the DFT representation for texture as (Kalidindi et al. 2009):

$$\widetilde{F}_{klm} = \sum_{b_1=0}^{N_{\varphi_1}-1} \sum_{b_2=0}^{N_{\varphi}-1} \sum_{b_3=0}^{N_{\varphi_2}-1} f_{b_1 b_2 b_3} \sin \varphi_{b_2} e^{\frac{-2\pi i k b_1}{N_{\varphi_1}}} e^{\frac{-2\pi i l b_2}{N_{\varphi}}} e^{\frac{-2\pi i m b_3}{N_{\varphi_2}}} \tag{18}$$

where $N_{\varphi_1} \text{x } N_{\varphi} \text{x } N_{\varphi_2}$ represents a uniform discretization of three-dimensional Bunge-Euler space, $(b_1, b_2, b_3)$ enumerate the grid points, and $f_{b_1 b_2 b_3}$ denotes the value of the ODF at the grid point.

In the Taylor model (Taylor 1938), the macroscopic deviatoric stress tensor, $\overline{\sigma}'$, is given by the volume averaged value of the local stress tensors in the constituent crystals of the polycrystalline aggregate. The volume-averaged value can be efficiently evaluated using Eqs. (10) and (18) and the orthogonal properties of the spectral representations. It can be shown that

$$\overline{\sigma}'_q = s \, \dot{\varepsilon}^m \, \frac{1}{N_{\varphi_1} \, N_{\varphi} N_{\varphi_2} N_{\theta}} \sum_k \sum_l \sum_m \sum_n \widetilde{F}_{klm} C_{klmn} e^{\frac{2\pi i n q}{N_{\theta}}} \tag{19}$$

where $\overline{\sigma}'$ denotes the volume averaged deviatoric stress tensor evaluated on a uniform grid in $\theta$, with the grid points enumerated by $q$. Equation (19) is used to establish the points on the yield surface corresponding to a particular principal frame of **D**. However, to establish the complete yield surface in the sample reference frame, one needs to explore the space of all possible principal frames. Thus, the space of all possible principal frames can be identified using a set of Euler angles, analogous to the Bunge-Euler angles used in the definition of the crystal lattice orientation. The spectral representations of the ODF and the stress functions do not implicitly assume any sample symmetry, and consequently, the yield surface can be constructed for any texture in the sample, without the need to invoke any simplifying assumptions
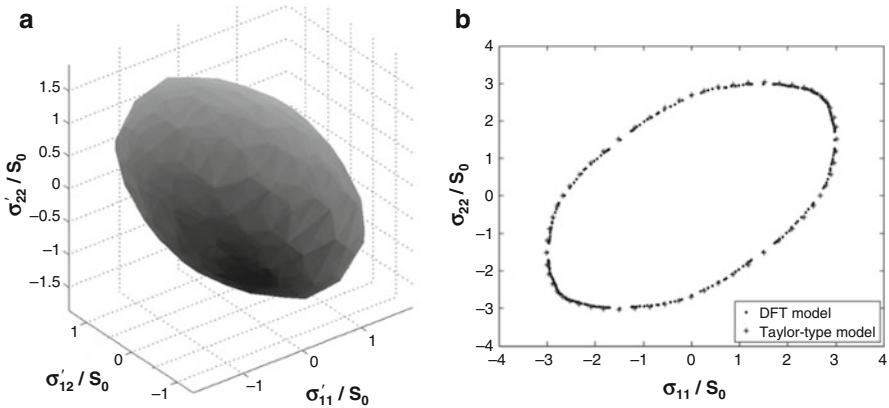
**Fig. 3** (**a**) Three-dimensional projection of the yield surface computed using the DFT-based spectral method for IF-steel with a random texture; (**b**) plots of the predicted $(\sigma_{11}, \sigma_{22})$ -yield locus for the same material comparing the spectral approach with the conventional Taylor-type model

of sample symmetry. It should be noted that since only a limited number of the $C_{klmn}$ transforms need to be accounted in evaluating Eq. (19), this leads to very efficient computations.

The results of the successfully implemented spectral method to construct the complete five-dimensional yield surface for BCC polycrystalline material are presented in Fig. 3. A set of 1000 discrete crystal orientations with random texture is used in the simulations. The time required for computing the entire five-dimensional yield surface (involving computations of the values of 7,200,000 stress tensors) was only 170 s. Figure 3a represents a selected projection of the five-dimensional yield surface computed here for IF-steel using 500 dominant DFTs. In order to check the accuracy of the yield surface, we compare in Fig. 3b the $(\sigma_{11}, \sigma_{22})$ section of the IF-steel yield surface computed using 500 dominant DFTs against the one computed using the conventional approach. The DFT-based computations are in excellent agreement with the conventional computations (Al-Harbi et al. 2010).

## 4.3    Plastic Closures Using Spectral Databases

Property closures delineate the complete set of all theoretically achievable combinations of selected effective anisotropic properties in a given material system and for a selected homogenization theory and are of great interest in optimizing the performance of engineering components. In general, these are very difficult to compute as they aim to map the complete space of theoretically feasible microstructures in the given material system into the property space of interest. Clearly, the availability of theoretically predicted closures for elastic-plastic properties of polycrystalline materials is of tremendous value in the design and development of new materials

with enhanced properties or performance characteristics. The complete space of all theoretically feasible microstructure distribution functions is identified as a microstructure hull, which for the first-order statistics in polycrystalline microstructures essentially constitutes a texture hull that is conveniently expressed using Fourier representations. First-order property closures in polycrystalline materials are essentially obtained by mapping the texture hulls into the selected property spaces.

The typical plastic properties of interest such as the tensile yield strength are defined in the stress space. The challenge consists in identifying an imposed deformation mode (i.e., stretching tensor, $\mathbf{D}$) that corresponds to the stress state. As an example, consider the computation of the uniaxial yield strength of the polycrystal. Since most crystal plasticity models (Kalidindi et al. 1992) take the imposed deformation as the input and predict the corresponding stress states, it becomes necessary to guess the deformation mode that would result in a uniaxial stress state. Without the assumption of orthorhombic sample symmetry, this search has to take place on at least a four-dimensional surface (equivalent to establishing the yield surface in five-dimensional stress space). However, with the assumption of orthorhombic sample symmetry, the search can be restricted to a single parameter space.

This case study demonstrates how the spectral methods can be utilized for the fast computation of the yield surface to produce a class of plastic property closures that include cubic-triclinic textures (Al-Harbi et al. 2010). It should be emphasized that the highly efficient computation of the yield surface using the spectral databases described in Sect 4.2 allows us to establish these plastic property closures. The methodology used here for building the first-order plastic closures follows the genetic-like algorithms described in Ref. (Knezevic et al. 2008b). The property combinations of interest were first evaluated for a set of crystal orientations that are uniformly distributed over the entire FZ. Then, weighted combinations of crystal orientations located on the boundary were used to expand the property closure. The process was repeated until the closure did not expand any more.

Since the property closures denote potential design spaces, it is important to ascertain how the cubic-triclinic closures expand the available design space. Therefore, in this case study, we computed the example plastic property closures based on the full-constraints Taylor model for both FCC copper and BCC IF-steel. Figure 4 shows the first-order closures delineating all of the feasible combinations of the normalized yield strengths in the sample $\mathbf{e}_1$ and $\mathbf{e}_2$ directions (i.e., $\sigma_{y1}$ and $\sigma_{y2}$) for copper and IF-steel computed assuming both orthorhombic and triclinic sample symmetries. Figure 4 clearly indicates that some combinations of $\sigma_{y1}$ and $\sigma_{y2}$ cannot be attained with the cubic-orthorhombic textures. Comparison of the closures in Fig. 4a and b reveals that the difference between the cubic-orthorhombic and the cubic-triclinic property closures is considerably larger for FCC metals compared to BCC metals. This observation is attributed to the availability of the higher number of slip systems in the BCC metals (48 slip systems in BCC compared to only 12 in the FCC crystals). The higher number of slip systems is expected to lower the degree of anisotropy in the response of the bcc metals and should therefore reduce the
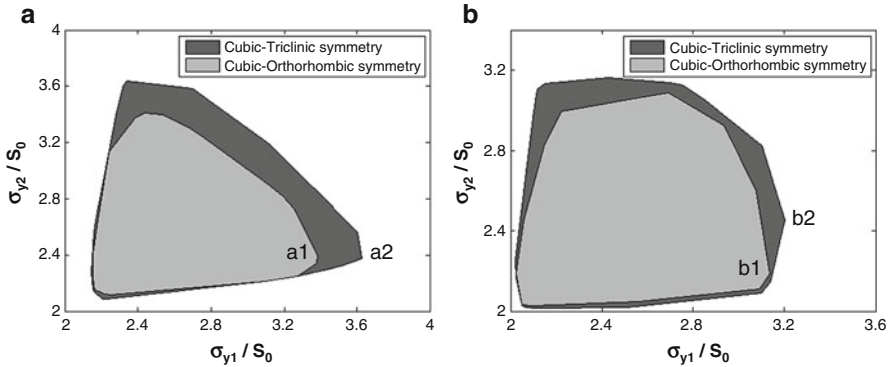
**Fig. 4** First-order cubic-triclinic and cubic-orthorhombic plastic closures for ($\sigma_{y1}/s_o$, $\sigma_{y2}/s_o$) computed using the DFT-based methods developed in this work. (**a**) OFHC Copper, (**b**) IF-steel

difference between cubic-orthorhombic and cubic-triclinic closures for these metals compared to the FCC metals.

## 4.4 Crystal Plasticity Based Forming Limit Diagram Predictions Using a Spectral Database

Accurate prediction of localized necking in thin metal sheets continues to represent a major challenge in the design of sheet metal forming of structural components. The onset of localized necking represents the ultimate deformation state that a stretched metal sheet can undergo, since this phenomenon is often precursor to material failure. The most common representation of this limit of material formability, under a variety of plane stress conditions (relevant to sheet forming), is through the concept of forming limit diagram (FLD). Since the experimental determination of FLDs has proven to be both expensive and difficult, considerable effort has been expended to develop reliable theoretical and/or numerical predictions. Such approaches require the use of a suitable constitutive framework to describe the evolution of the material state, in conjunction with an instability criterion, to predict the onset of plastic flow localization. A numerical tool has been developed recently (Gupta et al. 2018) for a fast and robust prediction of the forming limit diagrams (FLDs) for thin polycrystalline metal sheets using a Taylor-type (full constraints) crystal plasticity model. The occurrence of localized necking is predicted by the initial imperfection approach, which is based on the well-known Marciniak- Kuczynski (M-K) model (Marciniak et al. 1973; Marciniak and Kuczynski 1979). Recently, several multiscale models employing crystal plasticity constitutive descriptions have been coupled with the M–K analysis in order to predict localized necking and the associated FLDs. The solutions to the highly nonlinear crystal plasticity constitutive equations involved in these computations are known to be computationally very expensive. This presents a major impediment

to the wider adoption of crystal plasticity theories in the computation of FLDs. In this work, this limitation is addressed by using spectral database approach based on DFTs.

A rigid-viscoplastic framework was employed to model the mechanical behavior at the single crystal scale. Furthermore, we have restricted our attention to FCC crystals and employed Taylor-type scale-transition scheme in order to compute the overall mechanical behavior of the polycrystalline aggregate. This case study demonstrates the application of the DFT spectral database strategy to predict localized necking and the associated FLDs of thin metal sheets, represented as spectral crystal plasticity FLD (SCP-FLD). This new tool is validated by comparing the predicted SCP-FLDs with the predictions from conventional methods, represented as CP-FLD.

The M–K analysis is based on the assumption of a preexisting initial geometric imperfection in the form of a narrow band (also known as groove) across the width of the metal sheet. During the deformation, more plastic deformation occurs in the band, eventually leading to localized thinning in the band. FLD is predicted using M-K method in conjunction with spectral crystal plasticity by applying polycrystal model to two polycrystalline aggregates, representing the band and the homogeneous zone. For details of algorithm to predict FLD and other numerical aspects involved in this study, readers should refer to Gupta et al. (2018).

The spectral DFT database of one-degree grid is used with the SCP model for the FLD predictions (SCP-FLD). These predictions are compared in Fig. 5 with the FLDs predicted using conventional crystal plasticity model (CP-FLD). For both simulations, we have fixed the initial imperfection factor to 0.99. The polycrystals used in these simulations are made of 1000 FCC single crystals possessing initial random texture for Fig. 5a and initial textured material for Fig. 5b. In general, a textured material exhibits more anisotropic behavior. The initial texture corresponding to a textured material used in Fig. 5b corresponds to the texture predicted by a Taylor-type model after a plane strain compression (PSC) to a true strain of $-0.5$. The SCP parameters used for the FLD predictions are 500 dominant DFTs for deviatoric stress components and total shear rate and 2000 dominant DFTs for plastic spin components. By analyzing Fig. 5, we can easily see that the results obtained by both models (namely the SCP-FLD and CP-FLD) match very well and are indeed independent of the initial texture. The computational time (CPU time) required for predicting FLD reduces dramatically when using SCP model, resulting in about 96% savings.

## 5    Implementation of Spectral Databases in Finite Element Models

The case studies discussed so far used Taylor-type models, where a uniform deformation gradient is assumed for all grains. Crystal plasticity finite element models (CPFEM) would highly benefit from DFT databases, since computational time presents a significant constraint for CPFEM. This section explains how the

(a)                                                (b)
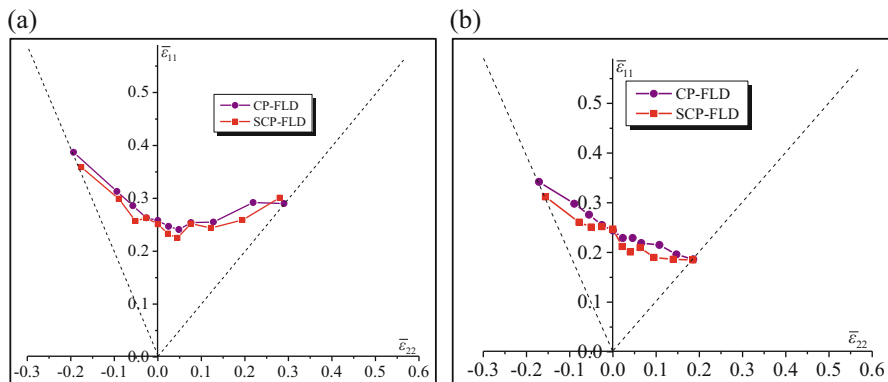


**Fig. 5** Predicted FLDs using conventional crystal plasticity model (CP-FLD) and spectral crystal plasticity model using DFT database (SCP-FLD) for the following hardening parameters: $h_0 = 500$ MPa, $s_s = 230$ MPa, $a = 2.8$, $s_0 = 50$ MPa and $m = 0.01$ and polycrystal consisting of 1000 single crystals corresponding to (**a**) random initial texture, (**b**) textured material (texture corresponding to plane strain compression to a true strain of $-0.5$ on the initially random texture using a Taylor-type model)

crystal plasticity DFT databases can be integrated with the commercial finite element package ABAQUS through a user materials subroutine (UMAT). To use the new spectral database scheme in the FE analysis, two tasks must be accomplished. Both of these developments have been discussed in Alharbi and Kalidindi (2015) and are summarized next.

## 5.1 Including Elastic Deformation in the DFT Database Approach

The crystal plasticity calculations using spectral databases need to be extended from rigid-viscoplastic behavior to elastic-viscoplastic deformation. Although the elastic deformation in most metals subjected to finite plastic deformation is indeed very small and can be neglected, it is essential to include elasticity for implementing crystal plasticity computations with most commercial FE codes. This is mainly because most FE simulation tools, such as ABAQUS, provide the total deformation gradient at each integration point as an input to the user-defined material constitutive response (through subroutines such as UMAT in ABAQUS) and expect to be returned the full stress tensor (not just the deviatoric stress tensor).

The following constitutive relations are used to include the elastic deformation with the spectral crystal plasticity approach described in the previous section:

$$\boldsymbol{\sigma}^{\nabla *} = \widetilde{\mathcal{L}} \, \mathbf{D}^* \tag{20}$$

where $\mathbf{D}^*$ is the elastic stretching tensor, $\widetilde{\mathcal{L}}$ is the fourth-rank elasticity tensor, and $\boldsymbol{\sigma}^{\nabla *}$ is the Jaumann rate of the Cauchy stress based on the axes that spin together

with the lattice and is defined as:

$$\boldsymbol{\sigma}^{\nabla *} = \dot{\boldsymbol{\sigma}} - \mathbf{W}^* \boldsymbol{\sigma} + \boldsymbol{\sigma} \mathbf{W}^* \tag{21}$$

In order to use the above relations, the total stretching tensor $\mathbf{D}$ (symmetric part of the velocity gradient tensor) needs to be decomposed into elastic and plastic parts. This decomposition must be accomplished such that the deviatoric stresses computed from both the crystal plasticity DFT databases (see Eq. (10), denoted here as $\boldsymbol{\sigma}'^{\mathrm{DFT}}(\mathbf{D}^p)$) and the above Jaumann rate relations (denoted as $\boldsymbol{\sigma}'^{\mathrm{Jmn}}(\mathbf{D}^*, \mathbf{W}^*, \Delta\mathrm{t})$) are equal to each other within an acceptable tolerance. The following modified Newton-Raphson scheme has been developed to accomplish this decomposition:

$$\left[\mathbf{D}'^*\right]_{n+1} = \left[\mathbf{D}'^*\right]_n - \lambda \ [\mathbf{J}]_n^{-1}[\mathbf{Err}]_n \tag{22}$$

$$\mathbf{Err} = \boldsymbol{\sigma}'^{\mathrm{DFT}}\left(\mathbf{D}^p\right) - \boldsymbol{\sigma}'^{\mathrm{Jmn}}\left(\mathbf{D}^*, \mathbf{W}^*, \Delta\mathrm{t}\right) \tag{23}$$

$$\mathbf{J} = \frac{\partial \mathbf{Err}}{\partial \mathbf{D}'^*} = -\frac{\partial \boldsymbol{\sigma}'^{\mathrm{DFT}}\left(\mathbf{D}^p\right)}{\partial \mathbf{D}^p} - \frac{\partial \boldsymbol{\sigma}'^{\mathrm{Jmn}}\left(\mathbf{D}^*, \mathbf{W}^*, \triangle\mathrm{t}\right)}{\partial \mathbf{D}^*} \tag{24}$$

The subscripts $n$ and $n + 1$ in Eq. (22) correspond to the estimates of $\mathbf{D}'^*$ at $n$ and $n + 1$ iterations, respectively, and the scalar parameter $\lambda$ is selected such that the magnitude of the step correction $\|\Delta\mathbf{D}'^*\| = \|[\mathbf{D}'^*]_{n+1} - [\mathbf{D}'^*]_n\| \leq \eta \, \varepsilon_{\mathrm{yield}}$, where $\varepsilon_{yield}$ denotes the magnitude of the total strain at yielding and $\eta$ is a numerical constant taken as 0.1.

## 5.2   Computation of the Jacobian

The implementation of UMAT in ABAQUS (ABAQUS 2010) requires the computation of the Jacobian defined as

$$\mathcal{J} = \frac{\partial \Delta \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\varepsilon}} \equiv \frac{\partial \boldsymbol{\sigma}}{\partial \mathbf{E}_t} \tag{25}$$

where $\Delta\boldsymbol{\sigma}$ and $\Delta\boldsymbol{\varepsilon}$ are the increments in the stress and strain tensors in a given time increment, respectively, and $\mathbf{E}_t$ is the relative strain tensor in the same time increment. The Jacobian matrix of Eq. (25) is used in the Newton-Raphson iterative method for revising the estimated displacements such that the corresponding stresses are likely to better satisfy the principle of virtual work at the end of the increment. It should be noted that the Jacobian matrix plays an important role in the rate of convergence of the solution to the global equilibrium equations but has no effect on the accuracy of the solution. The following analytical expression for the Jacobian is developed:
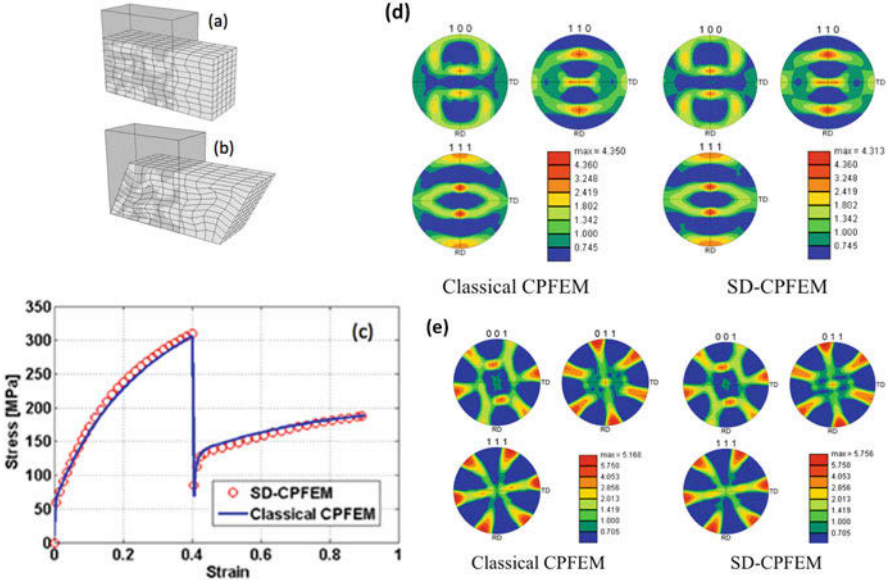
**Fig. 6** Comparison of the predicted stress-strain curves and texture from the SD-CPFEM against the corresponding results from the conventional CPFEM for plane strain compression followed by simple shear deformation of OFHC copper: (**a**) mesh after plane strain compression, (**b**) mesh after simple shear deformation, (**c**) stress-strain curves, (**d**) pole figures after plane strain compression, (**e**) pole figures after simple shear deformation

$$\mathcal{J} = \frac{\partial \boldsymbol{\sigma}}{\partial \mathbf{E}_t} = \frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{E}_t} + \mathbf{I} \otimes \frac{\partial p}{\partial \mathbf{E}_t} = \frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{D}'} \frac{\partial \mathbf{D}'}{\partial \mathbf{E}_t} + \mathbf{I} \otimes \frac{\partial p}{\partial \mathbf{E}_t} \tag{26}$$

with

$$\frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{D}'} = \left[ \mathbb{I} + \frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{D}^p} \frac{\partial \mathbf{D}'^*}{\partial \boldsymbol{\sigma}'} \right]^{-1} \frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{D}^p} \tag{27}$$

where $p$ denotes the pressure, and $\mathbf{I}$ and $\mathbb{I}$ are the second-rank and fourth-rank identity tensors, respectively. The term in Eq. (27) that requires long computations is $\frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{D}^p}$. It should be noted that, in any time step in the simulation, the term $\frac{\partial \boldsymbol{\sigma}'}{\partial \mathbf{D}^p}$ will be already calculated as a part of the iteration scheme to decompose the stretching tensor into elastic and plastic part (see Eq. (24)). Consequently, there is tremendous computational advantage in formulating the Jacobian computation as described in this section.

**Table 1** Elastic and plastic parameters of the OFHC copper used in the case study (Kalidindi et al. 1992)

| C$_{11}$ (MPa) | C$_{12}$ (MPa) | C$_{44}$ (MPa) | m | s$_o$ (MPa) | h$_o$ (MPa) | s$_s$ (MPa) | a |
|---|---|---|---|---|---|---|---|
| 168,400 | 121,400 | 75,400 | 0.01 | 16 | 180 | 148 | 2.25 |

# 6     Case Study of Spectral Database Application to CPFEM: Plane Strain Compression Followed by Simple Shear

In order to demonstrate the viability and computational advantages of the spectral CPFEM (SD-CPFEM), we present case study of simulations performed using spectral database and classical CPFEM models. The predictions from the two approaches are produced using the commercial FE package ABAQUS (ABAQUS 2010) and specially developed user material subroutines (described in Ref. (Alharbi and Kalidindi 2015)). To validate the SD-CPFEM for the case of nonproportional loading, we simulated a plane strain compression followed by simple shear of a FCC polycrystalline OFHC copper. The FE model consisted of 500 C3D8 elements with one crystal orientation per integration point. The initial texture is assumed to be random. Four thousand different crystal orientations with random texture were randomly assigned to each integration point inside each element. The first step in this simulation involved an imposed displacement boundary condition on the top surface of the model, which resulted in a 35% reduction in height corresponding to an axial true strain of about $-0.4$. All faces of the sample are constrained to remain planar in this step. In the second step, an imposed simple shear deformation is applied up to a shear strain of $\gamma = 0.5$ as shown in Fig. 6a and b. The elastic and plastic parameters of the OFHC copper used in this case study are shown in Table 1.

The stress-strain response from the SD-CPFEM is compared against the corresponding predictions from the classical CPFEM in Fig. 6c. The predicted textures from the two approaches after each deformation step are shown in Fig. 6 (d) and (e). It can be seen that the predictions from the spectral database approach matched very well with the corresponding predictions from the classical CPFEM at a dramatically reduced computation cost. This prediction took 6380 s for the classical CPFEM and only 527 s for the SD-CPFEM when using 500 DFTs for the stress, the shearing rate, and the lattice spin components. Examples of other case studies such as the simulation of extrusion and rolling process using the spectral database approach can be found in recent literature (Zecevic et al. 2015b).

# 7     Conclusions and Future Outlook

In this chapter, readers were introduced to recently developed spectral approach to crystal plasticity framework, where important variables of interest from crystal plasticity computations can be accessed directly from precomputed spectral database

based on discrete Fourier transform (DFT). The generation of databases requires one-time high computational cost, and once generated they can be used for wide variety of applications as demonstrated by various case studies discussed in this chapter.

The chapter provided details of DFT database generation for the deviatoric stresses, the lattice spins, and the strain hardening rates using Taylor-type crystal plasticity model. These variables depend on the crystal lattice orientations and deformation mode angle and database built for one-degree grid spacing in this 4-D grid gives very accurate results. Trigonometric interpolation can be used to obtain values for points other than grid points. The case studies using spectral database for FCC and BCC materials of Taylor-type simulations, computation of yield surface, property closures, and forming limit diagrams revealed tremendous savings in the computational time compared to conventional approaches based on crystal plasticity. The extension of the spectral database framework to do CPFEM simulations was also discussed with a case study that also shows a remarkable reduction in the computational time.

The computational benefit of spectral DFT database in speeding up crystal plasticity calculations is clearly evident from case studies discussed here. Further, it has been demonstrated in (Knezevic and Savage 2014) that there can be a further gain in the computational speed of doing crystal plasticity computations using spectral DFT database, if the computations are performed on a specialized computer hardware that utilizes graphics-processing units (GPU). Therefore, there is clear potential of further speeding up calculation using database if calculations are performed on advanced hybrid CPU–GPU computational platforms. The spectral database framework presented in this chapter will also be helpful to the manufacturing industry for performing multiscale simulations of forming operations where computational speed of crystal plasticity models is usually the bottleneck. Proper utilization of these toolsets can lead to accelerated insertion of new and improved materials into practice saving millions of dollars.

# References

ABAQUS (2010) Providence, RI, USA. © Dassault Systèmes Simulia Corp

Adams BL, Kalidindi S, Fullwood DT (2013) Microstructure-sensitive design for performance optimization. Butterworth-Heinemann, Waltham

Adzima F, Balan T, Manach PY, Bonnet N, Tabourot L (2017) Crystal plasticity and phenomenological approaches for the simulation of deformation behavior in thin copper alloy sheets. Int J Plast 94:171–191

Alharbi HF, Kalidindi SR (2015) Crystal plasticity finite element simulations using a database of discrete Fourier transforms. Int J Plast 66:71–84

Al-Harbi HF, Knezevic M, Kalidindi SR (2010) Spectral approaches for the fast computation of yield surfaces and first-order plastic property closures for polycrystalline materials with cubic-triclinic textures. Comput Mater Contin 15(2):153–172

Asaro RJ, Needleman A (1985) Overview no. 42 texture development and strain hardening in rate dependent polycrystals. Acta Metall 33(6):923–953

Bachu V, Kalidindi SR (1998) On the accuracy of the predictions of texture evolution by the finite element technique for fcc polycrystals. Mater Sci Eng A 257(1):108–117

Barlat F (2007) Constitutive modeling for metals. In: Advanced methods in material forming, pp 1–18, Springer, Berlin/Heidelberg

Bronkhorst CA, Kalidindi SR, Anand L (1992) Polycrystalline plasticity and the evolution of crystallographic texture in FCC metals. Philos Trans R Soc London, Ser A 341: 443–477

Bunge HJ, Esling C (1984) Texture development by plastic deformation. Scr Metall 18(3): 191–195

Callister WD Jr (2007) Materials science and engineering: an introduction. Wiley, New York

Chockalingam K, Tonks MR, Hales JD, Gaston DR, Millett PC, Zhang L (2013) Crystal plasticity with Jacobian-free Newton–Krylov. Comput Mech:1–11

Gupta A, Bettaieb MB, Abed-Meraim F, Kalidindi SR (2018) Computationally efficient predictions of crystal-plasticity based forming limit diagrams using a spectral database. Int J Plast 103: 168–187

Kalidindi SR, Anand L (1994) Macroscopic shape change and evolution of crystallographic texture in pre-textured FCC metals. J Mech Phys Solids 42(3):459–490

Kalidindi SR, Duvvuru HK (2005) Spectral methods for capturing crystallographic texture evolution during large plastic strains in metals. Acta Mater 53(13):3613–3623

Kalidindi SR, Schoenfeld SE (2000) On the prediction of yield surfaces by the crystal plasticity models for fcc polycrystals. Mater Sci Eng A 293(1–2):120–129

Kalidindi SR, Bronkhorst CA, Anand L (1992) Crystallographic texture evolution in bulk deformation processing of fcc metals. J Mech Phys Solids 40(3):537–569

Kalidindi SR, Bhattacharyya A, Doherty RD (2004) Detailed analysis of grain-scale plastic deformation in columnar polycrystalline aluminum using orientation image mapping and crystal plasticicty models. Proc Math Phys Eng Sci 460(2047):1935–1956. http://www.jstor.org/stable/4143061

Kalidindi SR, Duvvuru HK, Knezevic M (2006) Spectral calibration of crystal plasticity models. Acta Mater 54(7):1795–1804

Kalidindi SR, Knezevic M, Niezgoda S, Shaffer J (2009) Representation of the orientation distribution function and computation of first-order elastic properties closures using discrete Fourier transforms. Acta Mater 57(13):3916–3923

Knezevic M, Kalidindi SR (2017) Crystal plasticity modeling of microstructure evolution and mechanical fields during processing of metals using spectral databases. JOM 69(5):830–838

Knezevic M, Savage DJ (2014) A high-performance computational framework for fast crystal plasticity simulations. Comput Mater Sci 83:101–106

Knezevic M, Kalidindi SR, Fullwood D (2008a) Computationally efficient database and spectral interpolation for fully plastic Taylor-type crystal plasticity calculations of face-centered cubic polycrystals. Int J Plast 24(7):1264–1276

Knezevic M, Kalidindi SR, Mishra RK (2008b) Delineation of first-order closures for plastic properties requiring explicit consideration of strain hardening and crystallographic texture evolution. Int J Plast 24(2):327–342

Knezevic M, Al-Harbi HF, Kalidindi SR (2009) Crystal plasticity simulations using discrete Fourier transforms. Acta Mater 57(6):1777–1784

Kocks UF, Mecking H (2003) Physics and phenomenology of strain hardening: the FCC case. Prog Mater Sci 48(3):171–273

Lebensohn RA (2001) N-site modeling of a 3D viscoplastic polycrystal using fast Fourier transform. Acta Mater 49(14):2723–2737

Lebensohn RA, Liu Y, Ponte Castañeda P (2004) On the accuracy of the self-consistent approximation for polycrystals: comparison with full-field numerical simulations. Acta Mater 52(18):5347–5361

Lebensohn RA, Tome CN, Castaneda PP (2007) Self-consistent modelling of the mechanical behaviour of viscoplastic polycrystals incorporating intragranular field fluctuations. Philos Mag 87(28):4287–4322

Lebensohn RA, Brenner R, Castelnau O, Rollett AD (2008) Orientation image-based micromechanical modelling of subgrain texture evolution in polycrystalline copper. Acta Mater 56(15):3914–3926

Lubliner J (2008) Plasticity theory. Dover, New York

Marciniak Z, Kuczynski K (1979) Forming limit curve for bending processes. Int J Mech Sci 21(10):609–621

Marciniak Z, Kuczynski K, Pokora T (1973) Influence of plastic properties of a material on forming limit diagram for sheet-metal in tension. Int J Mech Sci 15(10):789–800

Molinari A, Canova GR, Ahzi S (1987) A self consistent approach of the large deformation polycrystal viscoplasticity. Acta Metall 35(12):2983–2994

Needleman A, Asaro RJ, Lemonds J, Peirce D (1985) Finite element analysis of crystalline solids. Comput Methods Appl Mech Eng 52(1–3):689–708

Pan J, Rice JR (1983) Rate sensitivity of plastic-flow and implications for yield-surface vertices. Int J Solids Struct 19(11):973–987

Raabe D, Wang Y, Roters F (2005) Crystal plasticity simulation study on the influence of texture on earing in steel. Comput Mater Sci 34(3):221–234

Roters F, Eisenlohr P, Hantcherli L, Tjahjanto DD, Bieler TR, Raabe D (2010) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. Acta Mater 58(4):1152–1211

Taylor GI (1938) Plastic strain in metals. J Inst Met 62:307–324

Van Houtte P (1994) Application of plastic potentials to strain rate sensitive and insensitive anisotropic materials. Int J Plast 10(7):719–748

Van Houtte P, Li S, Seefeldt M, Delannay L (2005) Deformation texture prediction: from the Taylor model to the advanced Lamel model. Int J Plast 21(3):589–624

Zecevic M, McCabe RJ, Knezevic M (2015a) Spectral database solutions to elasto-viscoplasticity within finite elements: application to a cobalt-based FCC superalloy. Int J Plast 70:151–165

Zecevic M, McCabe RJ, Knezevic M (2015b) A new implementation of the spectral crytsal plasticity framework in implicit finite elements. Mech Mater 84:114–126

# Advances in Computational Mechanics to Address Challenges in Crystal Plasticity FEM

# 79

## Somnath Ghosh

## Contents

**Abstract**

Image-based computational models are essential for predicting microstructure-property relationships. Crystal plasticity finite element models, or CPFEM, constitute a major part of these computational models. These models generally adopt conventional finite element analysis tools such as available commercial codes.

S. Ghosh (✉)
JHU Center for Integrated Structure-Materials Modeling and Simulation (CISMMS), Air Force Center of Excellence on Integrated Materials Modeling (CEIMM), US Association of Computational Mechanics (USACM), Department of Civil Engineering, Departments of Mechanical Engineering and Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA
e-mail: sghosh20@jhu.edu

However, they face severe challenges when modeling complex microstructures undergoing extreme phenomena. This chapter examines a few challenges of conventional CPFEM and proposes remedies through advanced methods of computational mechanics. The methods discussed include element stabilization, multi-time-domain subcycling, and efficiency enhancement through adaptivity. It demonstrates the need for such numerical advances and the advantages gained. It provides motivation for looking beyond the available tools and making fundamental advances in field of computational mechanics that can benefit predictive modeling.

# 1    Introduction

Structural materials, like metals and alloys, are often characterized by microstructural heterogeneities in the form of nonuniform grain distributions, multi-colony sites, or polyphase aggregates in a polycrystalline ensemble. Morphological and crystallographic characteristics of microstructures strongly govern their mechanical behavior and failure response. For example, disparities in grain size, crystallographic orientations, micro-texture, and slip system resistance cause large stress concentrations and grain boundary dislocation pileups, leading to localized plastic flow and crack nucleation in Ti alloys (Sinha et al. 2006). Robust predictive models of deformation and failure, incorporating microstructural features and physical mechanisms, are necessary for effective material simulation and design.

The recent years have seen a surge in the use of image-based computational micro-mechanics models for predicting microstructure-property relationships. For polycrystalline microstructures undergoing large plastic deformation, image-based computational modeling entails determining micromechanical solution fields in statistically equivalent representative volume elements (SERVEs) of the material microstructure by executing computational methods like the finite element method (FEM), boundary element method (BEM), or fast Fourier transform (FFT) method. A SERVE is an optimally small computational domain that is created to capture the statistics of characteristic morphological and crystallographic variables in experimentally obtained electron backscattered diffraction (EBSD) or scanning electron microscopy (SEM) images. Methods of generating SERVEs are discussed in a previous chapter (by Ghosh and Groeber) of this handbook. The SERVE creation is generally followed by conforming mesh generation through discretization of the microstructure into simple geometric elements. Figure 1 shows a 3D SERVE constructed from a 2D EBSD scan of a Ti-7Al specimen and strain evolution in the image-based crystal plasticity finite element analysis.

A majority of the computational models for polycrystalline microstructures undergoing large plastic deformation implement crystal plasticity constitutive models. Crystal plasticity finite element models, e.g., in Pierce et al. (1983), Busso et al. (2000), Staroselsky and Anand (2003), Matous and Maniatty (2004), Bridier et al. (2009), Roters et al. (2010a, b), Hasija et al. (2003), Deka et al. (2006), Venkataramani et al. (2007), Anahid et al. (2011), Meissonnier et al. (2001), Dunne et al. (2012), and Kalidindi and Schoenfeld (2000), generally incorporate these
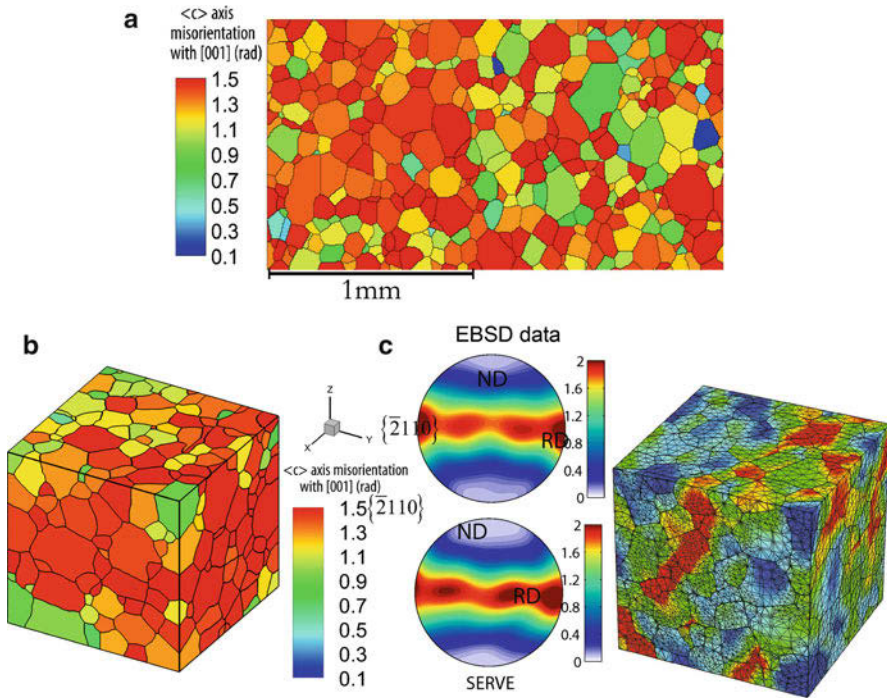
**Fig. 1** (**a**) EBSD scan of the alloy Ti-7Al, (**b**) 529-grain 3D SERVE of dimension 300 μm showing $< c >$-axis misorientation and pole figures comparing SERVE data to EBSD, (**c**) contour plot of plastic strain at 20% strain with an applied compressive strain rate of $10^4 \, \text{s}^{-1}$

constitutive models in conventional finite element analysis codes like ABAQUS, LS-DYNA, etc. for full field analysis of short- and long-range evolution of state variables. Image-based CPFEM, in which microstructural SERVEs are modeled for predicting complex deformation mechanisms including crack nucleation, twin propagation, etc., are discussed in Thomas et al. (2012), Shahba and Ghosh (2016), Cheng and Ghosh (2015, 2017), and Ozturk et al. (2017).

A number of challenges arise when using CPFEM for modeling deformation mechanisms in complex microstructures, especially those involving phenomena like localization, twinning, crack propagation, fatigue, etc. Special methods in advanced computational mechanics should be developed to overcome these challenges and render robust predictive tools. This chapter begins with a brief description of CPFEM formulations, crystal plasticity constitutive relations, and their implementation. Subsequently, it examines three challenges that commonly persist with conventional CPFEM and offers remedies for overcoming them.

- *Element Stabilization*: Plastic incompressibility causes volumetric locking of commonly used tetrahedral elements in CPFE analyses. Special element technology should be developed for stabilizing spurious modes in these elements.

- *Time-Domain Subcycling for Disparate Deformation Rates*: Modeling local-ization phenomena, e.g., discrete twin evolution, in CPFEM often has low computational efficiency due to very fine simulation time steps. This is a major bottleneck in predicting rapidly evolving twin bands. A multi-time-domain subcycling algorithm can improve computational efficiency through the introduction of complementary sub-domains with selective fine and coarse time stepping.
- *Enhanced Efficiency with Adaptive CPFEM*: Conventional 3D CPFEM with high-resolution mesh can be computationally prohibitive, especially with algo-rithms for solving complex constitutive models. Increased efficiency can com-promise accuracy due to the use of coarse mesh and simplified computational domains. Hierarchical adaptive methods in CPFEM are capable of providing a solution to this shortcoming.

## 2    Crystal Plasticity FE Formulation and Solution Methods

Finite deformation, crystal plasticity finite element models typically invoke an incremental solution method, where the time (or equivalent loading) domain is discretized into finite number of steps. In an updated Lagrangian formulation (Bathe 2006) for a body under quasi-static conditions, the principle of virtual work for an increment transcending discrete temporal points $t$ and $t + \Delta t$ is written as:

$$\int_{\Omega^t} \delta \mathbf{E}_t^{t+\Delta t} : \mathbf{S}_t^{t+\Delta t} \, d\Omega = R^{\text{ext}\, t+\Delta t} \tag{1}$$

where $\Omega^t \subset \mathcal{R}^3$ is the computational domain at time $t$ and $\mathbf{E}_t^{t+\Delta t}$ and $\mathbf{S}_t^{t+\Delta t}$ correspond to the Green-Lagrange strain and the second Piola-Kirchhoff (PK) stress tensors, respectively, with respect to the reference configuration at time $t$. The weak form at time $t + \Delta t$ requires the following relations:

Incremental displacement : $\mathbf{u}^{t+\Delta t} = \mathbf{u}^t + \Delta \mathbf{u}$ \hfill (2a)

Green-Lagrange strain tensor : $\mathbf{E}_t^{t+\Delta t} = \dfrac{1}{2} \left( \dfrac{\partial \Delta \mathbf{u}}{\partial \mathbf{x}^t} + \left( \dfrac{\partial \Delta \mathbf{u}}{\partial \mathbf{x}^t} \right)^T + \left( \dfrac{\partial \Delta \mathbf{u}}{\partial \mathbf{x}^t} \right)^T \dfrac{\partial \Delta \mathbf{u}}{\partial \mathbf{x}^t} \right)$

\hfill (2b)

Second Piola-Kirchhoff stress : $\mathbf{S}_t^{t+\Delta t} = J_t^{t+\Delta t} \left( \mathbf{F}_t^{t+\Delta t} \right)^{-1} \boldsymbol{\sigma}^{t+\Delta t} \left( \mathbf{F}_t^{t+\Delta t} \right)^{-T}$

\hfill (2c)

External virtual work : $R^{\text{ext}\, t+\Delta t} = \displaystyle\int_{\Omega^{t+\Delta t}} \delta \mathbf{u}^{t+\Delta t} \mathbf{b} \, d\Omega + \int_{\Gamma_\sigma^{t+\Delta t}} \delta \mathbf{u}^{t+\Delta t} \bar{\mathbf{t}} \, d\Gamma_\sigma$

\hfill (2d)

In these relations, $\boldsymbol{\sigma}$ is the Cauchy stress, $\mathbf{u}$ is the displacement, $\mathbf{b}$ is the body force per unit volume, $\delta\mathbf{u}$ is a displacement variation, $\mathbf{F}$ is the deformation gradient, and $J$ is its determinant or Jacobian. All quantities in Eqs. (2) are at time $t + \triangle t$ and referred to the configuration at time $t$. The second Piola-Kirchhoff (PK) stress and the Green-Lagrange strain are, respectively, decomposed as:

$$\mathbf{S}_t^{t+\triangle t} = \boldsymbol{\sigma}^t + \triangle\mathbf{S}^t \quad \text{and} \quad \mathbf{E}_t^{t+\triangle t} = \triangle\mathbf{E}^t = \triangle\mathbf{e}^t(\triangle\mathbf{u}) + \triangle\boldsymbol{\eta}^t(\triangle\mathbf{u}) \tag{3}$$

where $\triangle\mathbf{S}^t$ denotes the increment of second PK stress from time $t$ to $t + \triangle t$ and the Green-Lagrange strain tensor is decomposed into a linear part $\triangle\mathbf{e}^t(\triangle\mathbf{u}) = \frac{1}{2}\left[\left(\frac{\partial\triangle\mathbf{u}}{\partial\mathbf{x}^t}\right)^T + \frac{\partial\triangle\mathbf{u}}{\partial\mathbf{x}^t}\right]$ and a nonlinear part given as $\triangle\boldsymbol{\eta}^t(\triangle\mathbf{u}) = \frac{1}{2}\left(\frac{\partial\triangle\mathbf{u}}{\partial\mathbf{x}^t}\right)^T \frac{\partial\triangle\mathbf{u}}{\partial\mathbf{x}^t}$. Substituting the strain decomposition, the incremental crystal plasticity constitutive relation $\triangle\mathbf{S}^t = \mathbb{C}^t(\mathbf{u}) : \triangle\mathbf{E}^t(\mathbf{u}) \approx \mathbb{C}^t(\mathbf{u}) : \triangle\mathbf{e}^t(\mathbf{u})$, and the relation $\delta\mathbf{E}^t \approx \delta\mathbf{e}^t$ in Eq. (1), the linearized weak form is written as:

$$\int_{\Omega^t} \delta\mathbf{e} : \mathbb{C}^t(\mathbf{u}) \cdot \triangle\mathbf{e}(\mathbf{u})d\Omega + \int_{\Omega^t} \delta\boldsymbol{\eta} : \boldsymbol{\sigma}^t d\Omega = R^{\text{ext}\,t+\triangle t} - \int_{\Omega^t} \delta\mathbf{e} : \boldsymbol{\sigma}^t \, d\Omega \tag{4}$$

where $\mathbb{C}^t$ is the local material tangent stiffness matrix. The nonlinear weak form in Eq. (4) is solved by using an iterative scheme such as the Newton-Raphson method (Bathe 2006). In the $i$-th Newton-Raphson iteration, the spatially discretized linearized Eq. (4) is written as:

$$\mathbf{K}_t^i\mathbf{u} = \boldsymbol{b}_{t+\triangle t} - \mathbf{R}_t^i \quad \text{where} \quad \triangle\mathbf{u}^{i+1} = \triangle\mathbf{u}^i + \mathbf{u} = \triangle\mathbf{u}^i + \mathbf{u} \tag{5}$$

where $\mathbf{K}_t^i$ is the global tangent stiffness matrix, $\triangle\mathbf{u}^{i+1} = \triangle\mathbf{u}^i + \mathbf{u}$ is the displacement update, and $\boldsymbol{b}_{t+\triangle t} - \mathbf{R}_t^i$ is the residual force vector for every iteration. These are, respectively, expressed as:

$$\mathbf{K}_t^i = \int_{\Omega^t} \mathbf{B}^T\mathbb{C}^{t,i}\mathbf{B}d\Omega + \int_{\Omega^t} \mathbf{B}_{NL}^T\boldsymbol{\sigma}^{t,i}\mathbf{B}_{NL}d\Omega \, , \, \mathbf{R}_t^i = \int_{\Omega^t} \mathbf{B}^T\boldsymbol{\sigma}^{t,i}d\Omega$$

$$\boldsymbol{b}_{t+\triangle t} = \int_{\Omega^t} \mathbf{N}^T\mathbf{f}^{t+\triangle t} \, d\Omega + \int_{\Gamma^t} \mathbf{N}^T\boldsymbol{t}^{t+\triangle t} \, d\Gamma$$

where $\mathbb{C}^{t,i}$ is the elastoplastic tangent stiffness matrix in the $i$-th iteration; $\mathbf{B}$ and $\mathbf{B}_{NL}$ are the linear and nonlinear strain-displacement matrices, respectively; and $\mathbf{N}$ is the matrix of shape functions. The Newton-Raphson iterations continue till the residual $\boldsymbol{b}_{t+\triangle t} - \mathbf{R}_t^i$ reaches a predetermined tolerance.

## 2.1     Crystal Plasticity Constitutive Models

Crystal plasticity constitutive models account for dislocation glide on crystallographic slip systems. A significant body of work exists on micromechanical modeling using crystal plasticity models due to glide on slip systems, using, e.g., power law description (Pierce et al. 1983; Asaro and Needleman 1985), the thermally activated theory of plastic flow (Kocks et al. 1975), or the dislocation density-based models (Roters et al. 2010a, b). While computational mechanics advances are in general independent of the constitutive relations used, a crystal plasticity constitutive model for *hcp* materials developed in Cheng and Ghosh (2015, 2017) is summarized here.

### 2.1.1     Kinematic Relations, Flow Rule, and Slip System Resistances

The deformation gradient at time $t$ admits a multiplicative decomposition into elastic and plastic components as:

$$\mathbf{F}_0^t = \mathbf{F}^e \mathbf{F}^p \quad \text{where} \quad \det \mathbf{F}^e > 0 \quad \text{and} \quad \det \mathbf{F}^p = 1 \tag{6}$$

where $\mathbf{F}^e$ accounts for elastic stretching and rigid-body rotation of the crystal lattice, while $\mathbf{F}^p$ corresponds to the incompressible plastic flow due to slip. The second Piola-Kirchoff stress is expressed in terms of the elastic Green-Lagrange strain tensor $\mathbf{E}^e \left( = \frac{1}{2} \left( \mathbf{F}^{eT} \mathbf{F}^e - \mathbf{I} \right) \right)$ as:

$$\mathbf{S} = \mathbb{C}^e : \mathbf{E}^e \tag{7}$$

where $\mathbb{C}^e$ is a fourth-order anisotropic elasticity tensor. The evolution of plastic deformation is expressed in terms of plastic velocity gradient $\mathbf{L}^p$ as:

$$\mathbf{L}^p = \dot{\mathbf{F}}^p \mathbf{F}^{p-1} = \sum_{\alpha=1}^{N_{\text{slip}}} \dot{\gamma}^\alpha \mathbf{s}_0^\alpha \quad \text{where} \quad \mathbf{s}_0^\alpha = \mathbf{m}_0^\alpha \otimes \mathbf{n}_0^\alpha \tag{8}$$

Here $\dot{\gamma}^\alpha$ is the slip rate on a slip system $\alpha$, $N_{\text{slip}}$ is the total number of slip systems, and the Schmid tensor $\mathbf{s}_0^\alpha$ associated with $\alpha$-th slip system is expressed in terms of the slip direction $\mathbf{m}_0^\alpha$ and slip plane normal $\mathbf{n}_0^\alpha$ in the reference configuration. The dislocation glide-based slip rate on the slip system $\alpha$ is described in Pierce et al. (1983) and Cheng and Ghosh (2015, 2017) using a power law as:

$$\dot{\gamma}^\alpha = \dot{\gamma}_0^\alpha \left| \frac{\tau^\alpha - s_a^\alpha}{s_*^\alpha} \right|^{\frac{1}{m}} \text{sign}(\tau^\alpha - s_a^\alpha) \tag{9}$$

where $\dot{\gamma}_0^\alpha$ is a reference slip rate for slip system $\alpha$ and $m$ is the exponent representing strain-rate sensitivity. The resolved shear stress on slip system $\alpha$ is expressed as $\tau^\alpha = \mathbf{F}^{eT} \mathbf{F}^e \mathbf{S} : \mathbf{s}_0^\alpha$. The athermal shear resistance $s_a^\alpha$ is due to the interaction of

the stress field between parallel dislocation lines and from grain boundaries, and the thermal shear resistance $s_*^\alpha$ is due to local repelling obstacles, such as forest dislocations and dislocation jogs.

Two types of dislocations are considered in the evolution of athermal $s_a^\alpha$ and thermal $s_*^\alpha$ shear resistances. These are (i) statistically stored dislocations (SSDs) associated with the homogeneous components of plastic flow characterized by vanishing net Burgers vector and (ii) geometrically necessary dislocations (GNDs) corresponding to stored polarized dislocation densities. GND accumulation is necessary for accommodating crystal lattice curvatures in single crystal bending or near-grain boundaries of polycrystalline aggregates. The athermal and thermal hardening rates due to the evolution of SSDs are given as:

$$\dot{s}_{a,\text{SSD}}^\alpha = \sum_{\beta=1}^N h_a^{\alpha\beta} \ |\dot{\gamma}^\beta \sin(\mathbf{n_0}^\alpha, \mathbf{t_0}^\beta)| \ \ \text{and} \ \ \dot{s}_{*,\text{SSD}}^\alpha = \sum_{\beta=1}^N h_*^{\alpha\beta} \ |\dot{\gamma}^\beta \cos(\mathbf{n_0}^\alpha, \mathbf{t_0}^\beta)|$$

(10)

where $\mathbf{t_0}^\beta$ is the dislocation line tangent vector for edge dislocation on the slip plane $\beta$ and the coefficient matrices $h_a^{\alpha\beta}$ and $h_*^{\alpha\beta}$ represent the hardening of athermal and thermal shear resistances on the slip system $\alpha$ due to activity on slip system $\beta$, respectively. The GND contributions to the slip system hardening are derived from two sources, viz., (i) dislocation components $\rho_{\text{GND},P}^\alpha$ parallel to the slip plane $\alpha$ that cause hardening due to the athermal shear resistance $s_a^\alpha$ and (ii) forest dislocation components $\rho_{\text{GND},F}^\alpha$, which contribute to hardening due to thermal shear resistance $s_*^\alpha$ as:

$$s_{a,\text{GND}}^\alpha = c_1 G b \sqrt{\rho_{P,\text{GND}}^\alpha} \ \ \text{and} \ \ s_{*,\text{GND}}^\alpha = \frac{Q_{\text{slip}}}{c_2 c_3 b^2} \sqrt{\rho_{F,\text{GND}}^\alpha}$$

(11)

where $G$ is the shear modulus, $Q_{\text{slip}}^\alpha$ is the effective activation energy for dislocation slip, and $c_1$, $c_2$, $c_3$ are constants representing the passing stress, jump-width, and obstacle-width, respectively. GND accumulation can be measured in terms of the curl of the plastic deformation gradient per unit area in the reference configuration, which corresponds to the Nye's dislocation density tensor $\boldsymbol{\Lambda} = -(\boldsymbol{\nabla}_X \times \mathbf{F}^{p^T})^T$ (Anahid et al. 2011). The Nye's tensor is related to the GND density components on each slip system as (Cheng and Ghosh 2017):

$$\boldsymbol{\Lambda} = \sum_{\alpha=1}^{\text{nslip}} \rho_{\text{GND},s}^\alpha \mathbf{b_0}^\alpha \otimes \mathbf{m_0}^\alpha + \rho_{\text{GND},et}^\alpha \mathbf{b_0}^\alpha \otimes \mathbf{t_0}^\alpha + \rho_{\text{GND},en}^\alpha \mathbf{b_0}^\alpha \otimes \mathbf{n_0}^\alpha$$

(12)

where $\boldsymbol{\rho}_{\text{GND},s}$, $\boldsymbol{\rho}_{\text{GND},et}$, and $\boldsymbol{\rho}_{\text{GND},en}$ are the GND density components with screw, in-slip-plane edge, and normal-to-slip-plane edge characteristics, per unit volume in the reference configuration and $\mathbf{b_0}^\alpha$ is the Burgers vector for a slip system $\alpha$. For *hcp* crystals, there are more slip systems than the number of components in $\boldsymbol{\Lambda}$, and hence $\rho_{\text{GND},s}$, $\rho_{\text{GND},et}$, and $\rho_{\text{GND},en}$ are obtained by solving a constrained

minimization problem of minimizing the $L_2$ norm of the GND densities subject to the constraint Eq. (12) (Anahid et al. 2011). Screw and edge GND components $\rho_{\text{GND},s}$, $\rho_{\text{GND},et}$, and $\rho_{\text{GND},en}$ on each slip system contribute, respectively, to the parallel ($\rho_{\text{GND},P}^{\alpha}$) and forest ($\rho_{\text{GND},F}^{\alpha}$) components of GNDs. The total athermal and thermal resistances are expressed as the sum of a part related to the evolving dislocation structure and another related to defects such as Peierls resistance, impurities, and point defects, as:

$$s_a^{\alpha} = s_{a,0}^{\alpha} + \sqrt{(s_{a,\text{SSD}}^{\alpha})^2 + (s_{a,\text{GND}}^{\alpha})^2} \;\; \text{and} \;\; s_*^{\alpha} = s_{*,0}^{\alpha} + \sqrt{(s_{*,\text{SSD}}^{\alpha})^2 + (s_{*,\text{GND}}^{\alpha})^2} \tag{13}$$

where $s_{a,0}^{\alpha}$ and $s_{*,0}^{\alpha}$ are initial resistances, independent of the dislocation structure.

### 2.1.2 Numerical Implementation of Crystal Plasticity Constitutive Model

The numerical time integration algorithm integrates the set of coupled differential equations in the nonlocal constitutive model using the following steps (details provided in Cheng and Ghosh 2015, 2017):

*Step A*:   Update stresses, plastic strains, and all state variables, keeping the nonlocal GND density and twin variables fixed, with known values of deformation variables at time $t$, as well as a given deformation gradient $\mathbf{F}(t + \triangle t)$;

*Step B*:   Update the GND densities and their rates of hardening by evaluating $\nabla_X \times \mathbf{F}^{p^T}$, using values in adjacent elements.

An implicit update algorithm is implemented in step A with update in step B (Cheng and Ghosh 2015, 2017). For step A, the algorithm assumes that the primary unknown variable is the second Piola-Kirchhoff stress $\mathbf{S}$ and seeks its solution from a set of six nonlinear equations by a Newton-Raphson iterative solver while updating other deformation and state variables.

$$\mathbf{S}(t + \triangle t) = \mathbf{S}^{tr} - \sum_{\alpha=1}^{N_{\text{tot}}} \triangle \gamma^{\alpha} \left( \mathbf{S}(t + \triangle t), s_a^{\alpha}(t + \triangle t), s_*^{\alpha}(t + \triangle t) \right) \mathbf{B}^{\alpha} \tag{14}$$

where $N_{\text{tot}}$ is the total number of slip systems and $\mathbf{B}^{\alpha}$ is defined as:

$$\mathbf{B}^{\alpha} = \mathbb{C} : \left[ \frac{1}{2} \left( \mathbf{A}(t + \triangle t)\mathbf{s}_0^{\alpha} + \mathbf{s}_0^{\alpha}\mathbf{A}(t + \triangle t) \right) \right] \tag{15}$$

with $\mathbf{A}(t + \triangle t) = \mathbf{F}^{p^{-T}}(t)\mathbf{F}^T(t + \triangle t)\mathbf{F}(t + \triangle t)\mathbf{F}^{p^{-1}}(t)$. The trial stress is expressed as $\mathbf{S}^{tr} = \mathbb{C} : \frac{1}{2} (\mathbf{A}(t + \triangle t) - \mathbf{I})$. In the iterative solution of Eq. (14), a residual is defined for $i$-th iteration the as:

$$\mathbf{G}(\mathbf{S}^i(t + \triangle t)) = \mathbf{S}^i(t + \triangle t) - \mathbf{S}^{tr} + \sum_{\alpha} \triangle \gamma^{\alpha} \mathbf{B}^{\alpha} \tag{16}$$

the $i + 1$-th iteration update to $\mathbf{S}$ is obtained as:

$$\mathbf{S}^{i+1}(t + \triangle t) = \mathbf{S}^i(t + \triangle t) - \left.\frac{\partial \mathbf{G}}{\partial \mathbf{S}}\right|_i^{-1} \mathbf{G}(\mathbf{S}^i(t + \triangle t)) \tag{17}$$

In this update procedure for $\mathbf{S}(t + \triangle t)$, the slip system resistances are held fixed. After convergence, the increment of shear resistances from SSDs are updated using Eqs. (10). The next step B computes the GND densities and associated hardening increments. A nonlocal, invoking the element gradient operator is implemented to evaluate $\boldsymbol{\Lambda}$. The gradient operator is calculated from nodal values using element shape functions. A super-convergent patch recovery (SPR) method is implemented to evaluate the nodal values of $\mathbf{F}_p$ from Gauss quadrature points to evaluate $\nabla_X \times \mathbf{F}^{pT}$ at quadrature points. Evolution of plastic strain in image-based CPFEM simulations of a SERVE is shown in Fig. 1c.

## 3      Stabilization of Four-Noded Tetrahedral Elements for CPFEM

Polycrystalline microstructures of metals and alloys often have sharp and tortuous grain boundaries and multiple grain junctions. Discretization, conforming to these domains in 3D, is conveniently accomplished using four-noded tetrahedral or TET4 elements. However, it has been observed, e.g., in Matous and Maniatty (2004), Gee et al. (2009), and de Souza Neto et al. (2005), that these elements suffer from severe volumetric locking when simulating incompressible materials. Crystal plasticity constitutive models exhibit near isochoric plastic flow (det $\mathbf{F}^p = 1$) and large volumetric strains in TET4 elements. High stresses induce high spurious dilatational energy and considerable error due to element locking. This is attributed to low *incompressibility constraint ratio*, defined as the ratio of number of degrees of freedom to the number of incompressibility constraints in the FE mesh.

A number of methods have been proposed for the stabilization of volumetric locking in TET4 elements (Gee et al. 2009; de Souza Neto et al. 2005; Dohrmann et al. 2000; Nguyen-Thoi et al. 2009). A key idea in these methods is to associate nodal points with patches corresponding to an assembly of surrounding sub-elements and subsequently to selectively integrate the dilatational and deviatoric parts of the FE weak form over these patches. This process reduces the incompressibility constraint ratio. However, these methods are not effective for anisotropic CPFE formulations for which the stress or the elastoplastic tangent stiffness tensor cannot be split into volumetric and deviatoric components. A finite deformation element formulation with a F-bar patch method has been introduced in de Souza Neto et al. (2005), while mixed elements have been proposed in Matous and Maniatty (2004) with augmentation strain fields in conjunction with a linearly interpolated pressure field.

Based on developments in Cheng et al. (2016), this section discusses three locking-free stabilized finite element formulations for CPFE analysis. They include a node-based uniform strain (NUS) element, a locally integrated B-bar (LIB)-based element, and a F-bar patch (FP)-based element.

## 3.1 Node-Based Uniform Strain and Locally Integrated B-Bar Elements

The node-based uniform strain (NUS) formulation, introduced in Dohrmann et al. (2000), assigns a patch of sub-elements $\hat{\Omega}^s = \sum_{i=1}^{N^s} \alpha_s^i \, \Omega_s^i$ to a node $s$ in the finite element mesh as shown in Fig. 2a. Here $N^s$ is the number of TET4 elements attached to a node $s$, $\Omega_s^i$ is the volume contribution of the $i$-th element to the patch $\hat{\Omega}^s$, and $\alpha_s^i$ is a scalar weighting factor. Within each patch, the strain increment $\hat{\mathbf{e}}^s$ is uniform and obtained by weighted averaging from surrounding elements as:
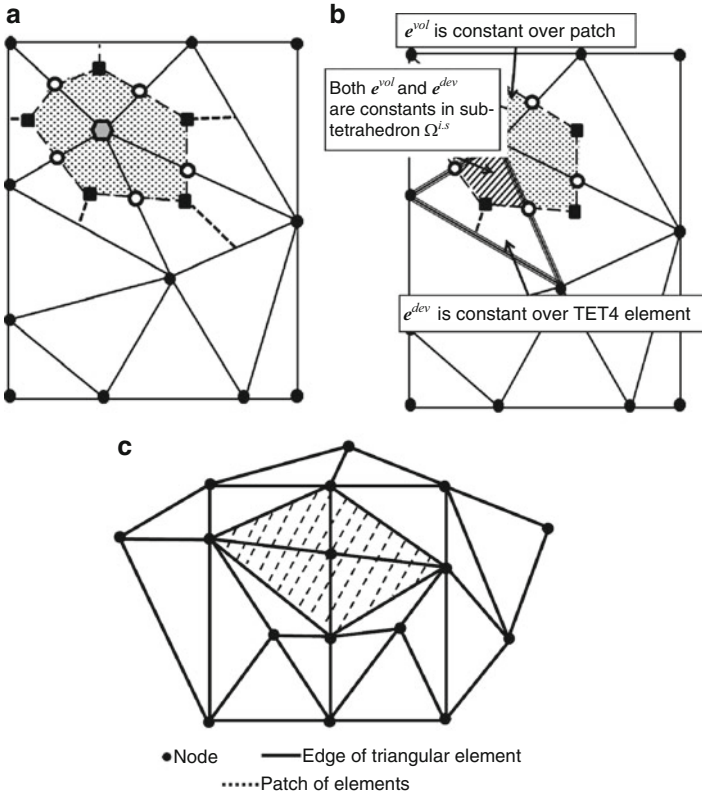


**Fig. 2** (**a**) 2D nodal patch in the NUS method, (**b**) nodal patch with strain distributions in the LIB method, and (**c**) patch of elements in the **F**-bar patch method

$$\hat{\mathbf{e}}^s = \sum_{i=1}^{N^s} w^i \mathbf{e}^{i,s} = \sum_{i=1}^{N^s} w^i \mathbf{B}^{i,t} \triangle \mathbf{q}^t = \hat{\mathbf{B}}^s \triangle \hat{\mathbf{q}}^s \tag{18}$$

$w^i$ is a relative volume-based weight for element $i$ and $\triangle \hat{\mathbf{q}}^s$ is the nodal displacement vector. $\hat{\mathbf{B}}^{s,t}$ is a strain-displacement matrix associated with the patch $s$, obtained by weighted assembling $\mathbf{B}^{i,t}$ from surrounding elements. The NUS formulation yields a constant tangent modulus and stress over the patch $s$, and one-point numerical integration is sufficient. However, this element can exhibit spurious zero or low-energy modes, causing large element distortion (Gee et al. 2009).

To overcome shortcomings of the NUS element for CPFE analysis with anisotropic elastoplastic stiffness matrix, a locally integrated B-bar (LIB) element is proposed in Cheng et al. (2016). This method alleviates volumetric locking without introducing spurious zero-energy modes. The strain increment is decomposed into volumetric and deviatoric parts by splitting the gradient matrix as $\mathbf{e} = \mathbf{e}^{\mathrm{vol}} + \mathbf{e}^{\mathrm{dev}} = \mathbf{B}^{\mathrm{vol}} \triangle \mathbf{q} + \mathbf{B}^{\mathrm{dev}} \triangle \mathbf{q}$. For reduced constraints, only the volumetric part of the strain increment $\mathbf{e}^{\mathrm{vol}}$ is assumed to be uniform inside the patch for each node $s$. The uniform volumetric strain increment $\hat{\mathbf{e}}^{s,\mathrm{vol}}$ is obtained as:

$$\hat{\mathbf{e}}^{s,\mathrm{vol}} = \sum_{i=1}^{N^s} w^i \mathbf{e}^{i,\mathrm{vol}} = \sum_{i=1}^{N^s} w^i \mathbf{B}^{i,\mathrm{vol}} \triangle \mathbf{q}^i = \bar{\mathbf{B}}^{s,\mathrm{vol}} \triangle \hat{\mathbf{q}}^s \tag{19}$$

$\bar{\mathbf{B}}^{s,\mathrm{vol}}$ is the volumetric part of the gradient matrix associated with a patch $s$. It is assembled from the $\mathbf{B}^{i,\mathrm{vol}}$ matrices of the surrounding elements with weights $w^i$. The deviatoric part of the strain increment $\mathbf{e}^{\mathrm{dev}}$ is constant in each of the contributing TET4 elements. This yields separate distributions of the volumetric and deviatoric strain increments over the domain, as illustrated in Fig. 2b. Each TET4 element is divided into four sub-domains of equal volume. Within each sub-domain, the volumetric and deviatoric parts of the strain increment are constant. The LIB element selectively reduces the volumetric strain components over the patch and keeps the deviatoric strain components unchanged.

## 3.2   F-Bar Patch-Based (FP) Element

The $\mathbf{F}$-bar patch (FP) formulation (de Souza Neto et al. 2005) has been shown to alleviate volumetric locking without introducing spurious zero-energy modes. The $\mathbf{F}$-bar patch method modifies the deformation gradient for weak enforcement of incompressibility in the element, rather than point-wise enforcement. The deformation gradient is decomposed into isochoric and volumetric components as:

$$\mathbf{F} = \mathbf{F}_{\mathrm{iso}} \mathbf{F}_{\mathrm{vol}} \quad \text{where} \quad \mathbf{F}_{\mathrm{iso}} = (\det \mathbf{F})^{-\frac{1}{3}} \mathbf{F} \text{ and } \mathbf{F}_{\mathrm{vol}} = (\det \mathbf{F})^{\frac{1}{3}} \mathbf{I} \tag{20}$$

Incompressibility in constitutive relations is enforced over a patch of elements, rather than in each element. This requires that elements in the mesh be assigned to non-overlapping patches as illustrated in Fig. 2c. For a patch $\mathscr{P}$ of a set of elements, with deformed and undeformed volumes $\Omega_{\text{patch}}^{t+\triangle t}$ and $\Omega_{\text{patch}}^{0}$, respectively, the deformation gradient in element $K \in \mathscr{P}$ is modified as:

$$\bar{\mathbf{F}}_K = \left[ \frac{\Omega_{\text{patch}}^{t+\triangle t}}{\Omega_{\text{patch}}^{0} \det \mathbf{F}_K} \right]^{\frac{1}{3}} \mathbf{F}_K \tag{21}$$

### 3.3    Performance of TET4 Element Stabilization in CPFEM

The performance of the locally integrated B-bar (LIB) element and the **F**-bar patch element with a patch size of 8 tetrahedrons (FP8) has been studied in Cheng et al. (2016) for CPFEM analysis of polycrystalline materials. Simulation results for a bicrystal compression test and a polycrystalline bending test of the *hcp* magnesium alloy *AZ*31 are compared with those for the standard TET4 element. The reference solution is for simulations with the B-bar stabilized eight-noded hexahedral element.

- **Bicrystal compression test**: A bicrystal of dimensions of $10 \times 10 \times 10\,\mu\text{m}$ is simulated under uniaxial compression using the standard TET4, LIB, and FP8 elements in CPFEM. Material constitutive models are given in Cheng and Ghosh (2015). The grain boundary is characterized by crystal orientations, which have Euler angles $[0°, 0°, 0°]$ and $[0°, 90°, 0°]$ in the $ZXZ$ convention for crystals 1 and 2, respectively, as shown in Fig. 3a. Displacement boundary conditions are applied on the top surface. A reference solution of the loading direction stress $\sigma_{zz}$ for CPFE analysis with hexahedral elements with **B**-bar stabilization is shown in Fig. 3b. Results using the standard, LIB and FP8 TET4 elements are shown in Fig. 3c, e, respectively. Non-smooth distribution of the local stress with high stress concentration is observed at the grain boundary using standard TET4 element, compared to other stabilized elements. The stress error is evaluated as the $L2$ norm of the difference with the reference solution as $\|e\|_{L2} = \frac{\left[ \int_{\Omega} \left( \sigma_{ij} - \sigma_{ij}^{ref} \right) \left( \sigma_{ij} - \sigma_{ij}^{ref} \right) d\Omega \right]^{\frac{1}{2}}}{\left( \int_{\Omega} \sigma_{ij} \sigma_{ij} d\Omega \right)^{\frac{1}{2}}}$. The corresponding error plots for different elements with increasing mesh densities are shown in Fig. 4a. The average convergence rate for LIB and FP8 elements is 0.75. For CPFE analysis, these elements exhibit similar results with much smaller errors compared to the standard TET4 element. The evolution of hydrostatic stress at the grain boundary with increasing strain is plotted in Fig. 4b. Unrealistically large stresses are observed with conventional TET4 elements, while the LIB and FP8 elements produce results that are consistent with the stabilized hexahedral element.
- **Micro-twin nucleation in polycrystalline magnesium alloy**: CPFE simulations are conducted for twin nucleation using a microstructural model of the
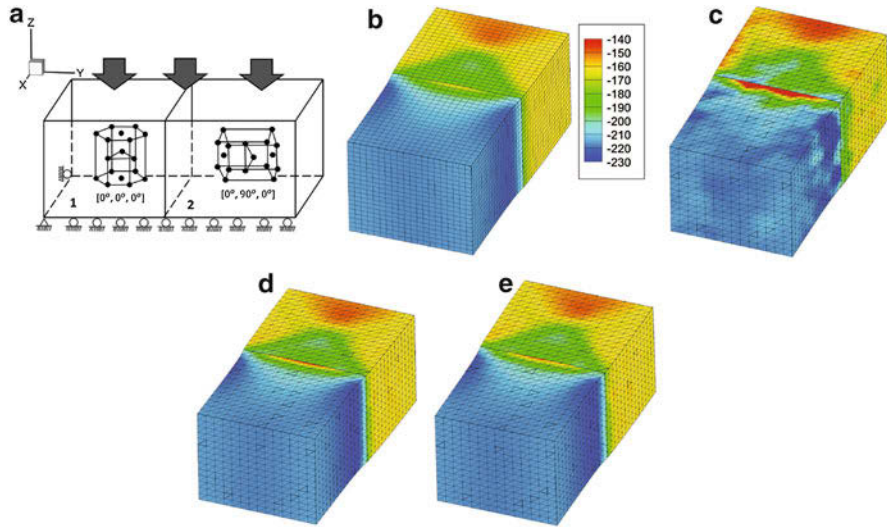
**Fig. 3** (**a**) Boundary conditions and crystallographic orientations in the bicrystal compression test; distribution of loading direction stress at 5% strain using simulation results of (**b**) eight-noded B-bar hexagonal elements with a mesh of 18,081 nodes, (**c**) standard TET4 element, (**d**) LIB element, and (**e**) FP8 element, all with a mesh of 11,862 nodes

polycrystalline Mg alloy AZ31 shown in Fig. 5a. The $40 \times 40 \times 40\,\mu$m statistically equivalent virtual microstructure with 103 grains of average size of $10\,\mu$m is developed from electron backscattered diffraction data. The microstructure is discretized into 113,425 TET4 elements. Displacement boundary conditions are applied at a rate of $0.004\,\mu$m/s on the two surfaces in Y-direction, which bend the microstructure about the X-axis on Y-Z plane. Details of the twin nucleation model are given in Sect. 4.1. The GND density contour plots in Fig. 5b, c show highest GND concentrations close to grain boundaries. The conventional TET4 element shows a much stiffer response in Fig. 6b compared to the other elements. The FP8 element shows a slightly lower level of locking than LIB elements due to a lower constraint ratio. The twin nucleation predictions in Fig. 6a show a much earlier twin nucleation time (97 s) with the TET4 elements in comparison with the LIB element (160 s) and FP8 element (180 s).

In summary, both the LIB and FP elements stabilize the local stresses and GND distributions in CPFE analyses and converge to the reference solution. The FP element is capable of providing slightly better results than the LIB element for an optimal patch size. A study on computational efficiency in Cheng et al. (2016) have shown that the FP element outperforms the LIB element with a considerably lower simulation time. From accuracy and efficiency considerations, the FP element is deemed more suitable for stabilized CPFE analysis.
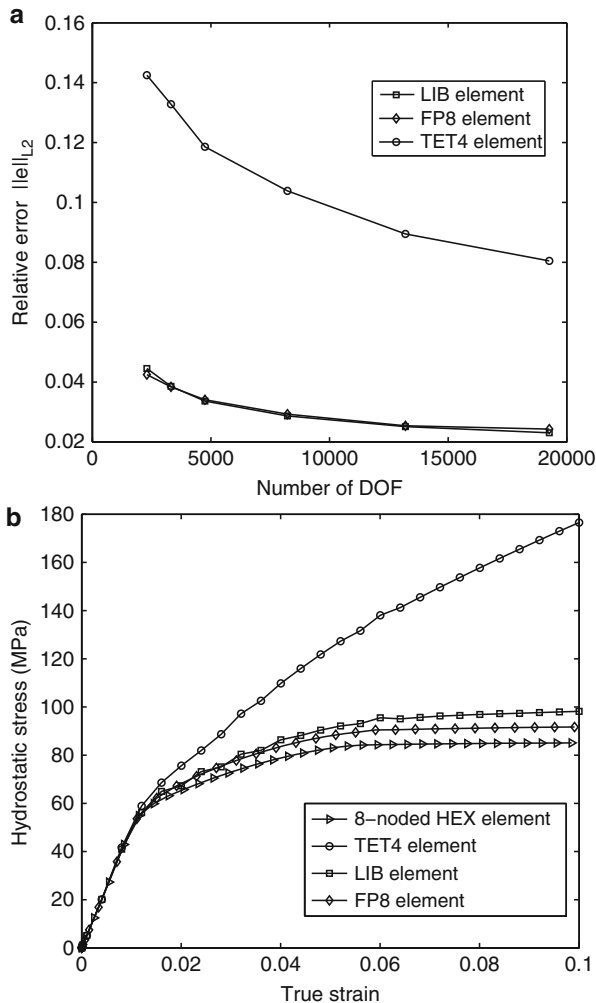
**Fig. 4** (**a**) Stress error plot with increasing degrees of freedom and (**b**) evolution of maximum hydrostatic stress with strain for different element formulations

# 4 Multi-Time-Domain Subcycling for Discrete Twin Evolution

Deformation twinning is a critical deformation mechanism that causes change in lattice orientations with localized deformation inside thin twin bands. Twinning can induce characteristic features like plastic anisotropy, tension-compression asymmetry, and local softening in the material response. Many crystal plasticity-based twinning models adopt a twin volume fraction approach that treats twin
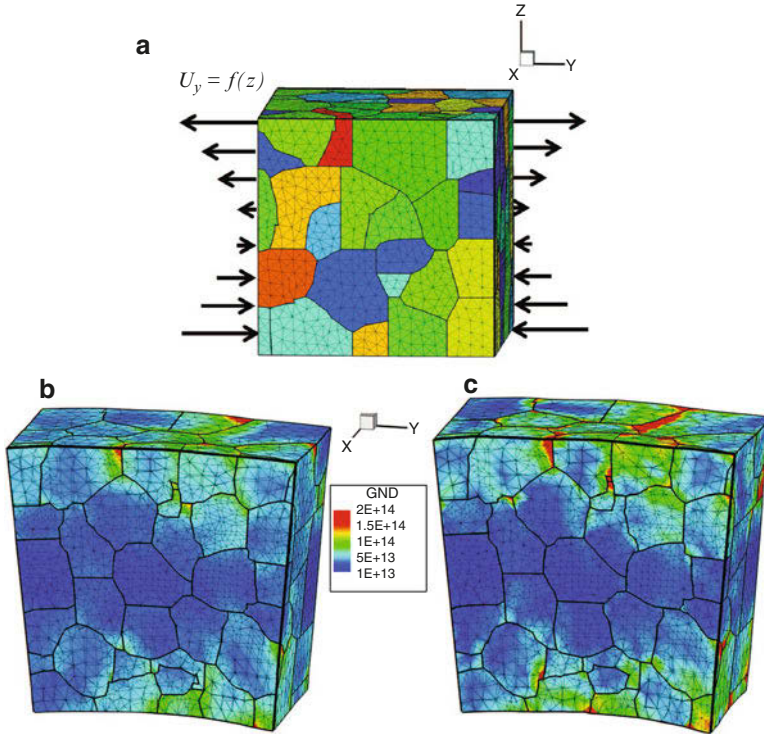
**Fig. 5** Schematic of the polycrystalline AZ31 SERVE showing the applied boundary conditions; GND densities distribution after 500s using: (**b**) TET4 and (**c**) FP8 elements

evolution in the same way as slip (Staroselsky and Anand 2003; Izadbakhsh et al. 2011; Zhang and Joshi 2012). These approaches do not account for deformation heterogeneity within discrete twins. Explicit twin formation models within the CPFE framework have been proposed based on phenomenological twin formation criteria and adaptive mesh generation methods in Abdolvand and Daymond (2013) and Knezevic et al. (2016). Explicit twinning models hold promise, provided the physics of twin nucleation, propagation, and interactions are correctly accounted for. The author has recently implemented an image-based crystal plasticity FE model with discrete twin evolution in Cheng and Ghosh (2017) and Cheng et al. (2018) to study deformation and twinning mechanisms in polycrystalline microstructures of Mg alloys.

A major difficulty with image-based CPFE simulations of polycrystalline microstructures delineating explicit twin formation is the high demands on computing time. This is attributed to the discrepant deformation rates between the domains of rapidly evolving twins and the surrounding crystalline matrix. The high rates of twin evolution in localized bands lead to numerical instability with the stiff nonlinear crystal plasticity constitutive equations, requiring very fine simulation
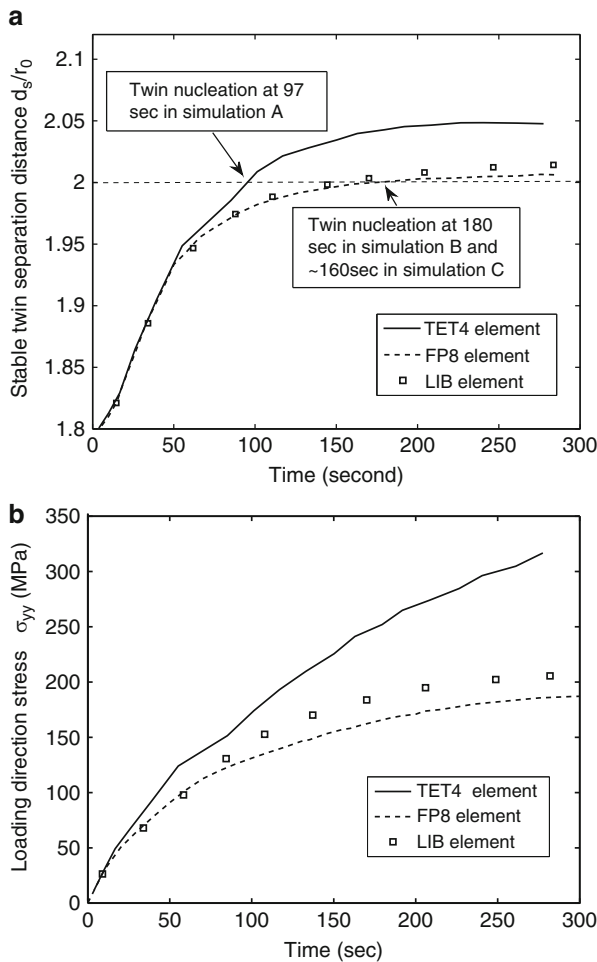
**Fig. 6** (**a**) Micro-twin dissociation distance as a function of loading time and (**b**) loading direction stress at a material point with loading time

time steps. A quantitative study on the critical time step size required for numerical stability for a CPFE simulation of a Mg alloy with evolving explicit twins has been conducted in Ghosh and Cheng (2018). The study shows that while 97% of elements require a minimum time step size of $\Delta t = 10$ s, only 3% of the elements located in localized twin bands require a significantly smaller time step of $\Delta t = 0.0391$ s. The reduction in time step size corresponds to a factor of $\sim$255. Requiring all elements in the computational domain to be integrated with the smaller step size can cause a huge loss of efficiency and could potentially be computationally intractable. While single time step methods to accelerate slip-based crystal plasticity models have been proposed, e.g., in Roters et al. (2010a), they are insufficient for twin evolution models.

## 4.1    Crystal Plasticity Constitutive Models with Twinning

The crystal plasticity constitutive models in Sect. 2.1 are extended in this section to account for both dislocation glide and twinning on crystallographic slip and twin systems. Twinning reorients the crystallographic lattice symmetrically by reflection across a mirror or twin plane in the reference configuration. In the twinned regions, plastic flow takes place by gliding of twin partial dislocations on twin planes, as well as by dislocation glide. The plastic velocity gradient in twinned region is subsequently expressed as:

$$\mathbf{L}^p = \dot{\mathbf{F}}^p \mathbf{F}^{p-1} = \sum_{\beta=1}^{N_{\text{twin}}} \dot{\gamma}_{tw}^{\beta} \mathbf{s}_{0,tw}^{\beta} + \sum_{\alpha=1}^{N_{\text{slip}}} \dot{\tilde{\gamma}}^{\alpha} \tilde{\mathbf{s}}_{0,\text{slip}}^{\alpha} \text{ where } \mathbf{s}_{0,tw}^{\beta} = \mathbf{m}_{0,tw}^{\beta} \otimes \mathbf{n}_{0,tw}^{\beta} \quad (22)$$

$\dot{\gamma}_{tw}^{\beta}$ is the shearing rate on a twin system $\beta$, $\dot{\tilde{\gamma}}^{\alpha}$ is the slip rate in the reorientated slip system $\alpha$, $N_{\text{twin}}$ is the number of twin systems, $\mathbf{m}_{0,tw}^{\beta}$ is the twin shearing direction vector, and $\mathbf{n}_{0,tw}^{\beta}$ is the twin plane normal in the reference configuration. Dislocation slip in the twinned volume occurs on a slip plane $\tilde{\mathbf{n}}_{0,\text{slip}}^{\alpha}$ in the direction $\tilde{\mathbf{m}}_{0,\text{slip}}^{\alpha}$, with mirror symmetry to the directions $\mathbf{n}_{0,\text{slip}}^{\alpha}$ and $\mathbf{m}_{0,\text{slip}}^{\alpha}$ in the matrix region. Twin nucleation and propagation models, developed in Cheng and Ghosh (2015, 2017) for *hcp* materials, are summarized here.

- *Twin nucleation*: The twin nucleation model is based on the elastic dislocation theory of twin nucleation by non-planar dissociation of a sessile pyramidal $\langle c + a \rangle$ dislocation. Three simultaneous conditions should be satisfied for the dissociation process to spontaneously occur and form a stable twin nucleus. They are (i) dissociation condition, $E_{\text{ini}} \geq E_{tw}(d = 0) + E_r$; (ii) irreversibility condition, $E_{\text{ini}} > E_F (d = d_s, \tau_{tw})$; and (iii) reliability condition, $d_s > 2r_0$. Here $E_{\text{ini}}$ is the initial system energy prior to dissociation, corresponding to the dislocation line self-energy of sessile pyramidal $\langle c + a \rangle$ dislocations, $E_{tw}$ is the self-energy of the twinning dislocation loop, $E_r$ is the self-energy of stair-rod dislocations, and $E_F$ is the post-dissociation total system energy. The distance $d$ is the separation distance between two partial dislocations and $d_s$ is the stable separation distance. Detailed description of the various energies and other critical parameters in the twin nucleation model are provided in Cheng and Ghosh (2015, 2017).
- *Explicit twin propagation*: Twin propagation involves two motions, viz., twin elongation by rapid gliding of twin partial dislocations on twin planes and twin thickening by migrating twin boundaries from the current twin planes to every other $\{10\bar{1}2\}$ twin plane (Keshavarz and Ghosh 2013, 2015). Gliding of twin partial dislocations occurs by a mixed shear-shuffle process, for which twin propagation is deemed as thermal activation process. The velocity of twin partial dislocation on a twin plane is expressed as:

$$v_{\text{glide}} = f_{\text{shuffle}}\lambda_{\text{shear}} \left[ \exp\left( -\frac{\triangle F - \tau A_P b_{tw}}{K_B T} \right) \right] \tag{23}$$

where $f_{\text{shuffle}}$ is the shuffling frequency, $\lambda_{\text{shear}}$ is the shear distance, $b_{tw}$ represents the Burgers vector on a twin system, $\tau$ is the effective resolved shear stress on the twin plane, and $A_P$ is the shearing area. The term $\exp\left( -\frac{\triangle F - \tau A_P b_{tw}}{K_B T} \right)$ is the probability of gliding with an internal energy barrier $\triangle F$. The twin boundary migrates when the twin partial dislocation glides on the adjacent twin plane.

A stimulated slip model is used to model thickening of twins. It assumes the existence of immobile lattice dislocations, which penetrate multiple twin planes. The velocity of twin partial dislocations crossing twin planes has been derived as:

$$v_{\text{thick}} = \frac{d_{tw}}{\triangle t_{tw}} = d_{tw} P_{\text{promoter}} \rho_{\text{tot}} l_{tw} v_{\text{glide}} \tag{24}$$

where $d_{tw}$ denotes the distance between twin planes, $\triangle t_{tw}$ is the average time required to meet a promoter, $P_{\text{promoter}}$ is the fraction of dislocations that act as promoters, $\rho_{\text{tot}}$ is the total dislocation density, and $v_{\text{glide}}$ and $l_{tw}$ are, respectively, the velocity and length of moving dislocations. The time-averaged, twin system plastic shear rate in Eq. (22) due to twinning is obtained from the Orowan equation as:

$$\dot{\gamma}_{tw} = \rho_{tw} b_{tw} v_{\text{glide}} = \dot{\gamma}_{0,tw} \exp\left( -\frac{\triangle F - \tau A_P b_{tw}}{K_B T} \right) \tag{25}$$

where $\dot{\gamma}_{0,tw} = \rho_{tw} b_{tw} f_{\text{shuffle}} \lambda_{\text{shear}}$ and $\rho_{tw}$ is the density of twin partial dislocations. For $\{10\bar{1}2\}$ twin in Mg, $\gamma_{tw}^{\max} = 0.1289$. In simulations, the shear on a twin system quickly reaches this upper bound once an integration point is twinned.

The complex interaction between dislocations and twin boundaries is modeled by the evolution of a slip/twin system resistance in which Eq. (25) is reduced to a conventional shear resistance-based power law model as:

$$\dot{\gamma}_{tw} = \dot{\gamma}_{0,tw} \left| \frac{\tau}{s_{tw}} \right|^{\frac{\triangle F}{K_B T}} \text{sign}(\tau) \tag{26}$$

For a twin system $\alpha$, the rate of shear resistance is expressed as $\dot{s}_{tw}^{\alpha} = \sum_{\beta=1}^{N_{\text{slip}}} h^{\alpha\beta} |\dot{\gamma}^{\beta}|$, where the hardening matrix $h^{\alpha\beta}$ quantifies hardening due to dislocation slip in the twinned regions. Implementation of twin evolution in the CPFEM framework is discussed in Cheng and Ghosh (2017) and Cheng et al. (2018).

## 4.2    Adaptive Subcycling for Accelerated CPFEM

An adaptive multi-time-domain subcycling algorithm is developed in Cheng and Ghosh (2017) and Ghosh and Cheng (2018) to avert the low efficiency due to minimum critical time step requirements, by activating a differential temporal resolution in the computational domain. The algorithm partitions the microstructural domain into sub-domains that are classified as critical (high strain rate) and noncritical (low-strain rate). For optimal efficiency, time integration in each sub-domain is conducted with its own independent time step, as determined from stability and accuracy criteria. For a twinned microstructure, regions of twin bands are solved with fine time steps, while the remaining regions use coarse time steps. A schematic layout of the algorithm is shown in Fig. 7. With a known state at time $t$, the integration algorithm for the time increment $t$ to $t + \triangle t$ solves the noncritical sub-domain problem using the coarsest possible time increment $\triangle t$ and the critical sub-domain problem using fine time steps $\triangle \tau \ll \triangle t$. To achieve global equilibrium for the computational domain, the different sub-domains are coupled, and residuals at the interfaces of discrepant time steps in the assembled sub-domains are minimized using a predictor-corrector scheme. The subcycling algorithm evaluates displacement correctors by equilibrating nodal residual forces since displacement fields at nodal points of adjacent sub-domains will not satisfy compatibility in general. Decomposition of the computational domain necessitates the evaluation of critical time steps for each element. With known state variables at time $t$, the critical time increment in each element is estimated from convergence of time integration of the constitutive model. If the time integration fails to converge, a scaled reduction of the original time increment is made. The essential steps in a staggered algorithm for evolving twins are given next and detailed in Cheng and Ghosh (2017) and Ghosh and Cheng (2018).
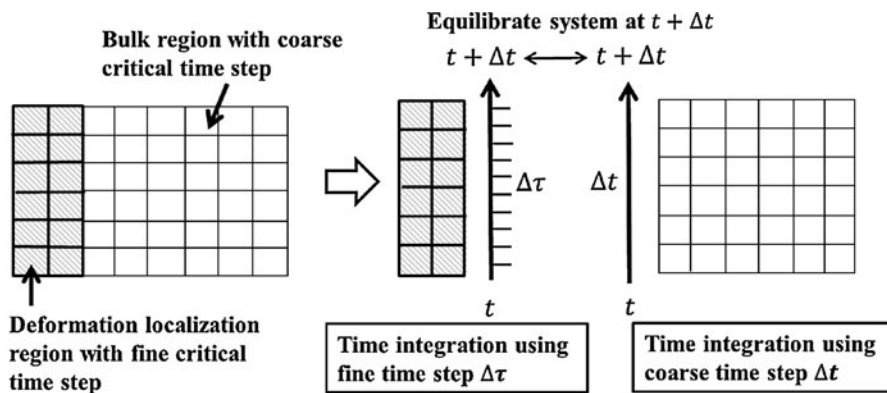


**Fig. 7** Schematic of the subcycling algorithm showing partitioning and equilibrating domains
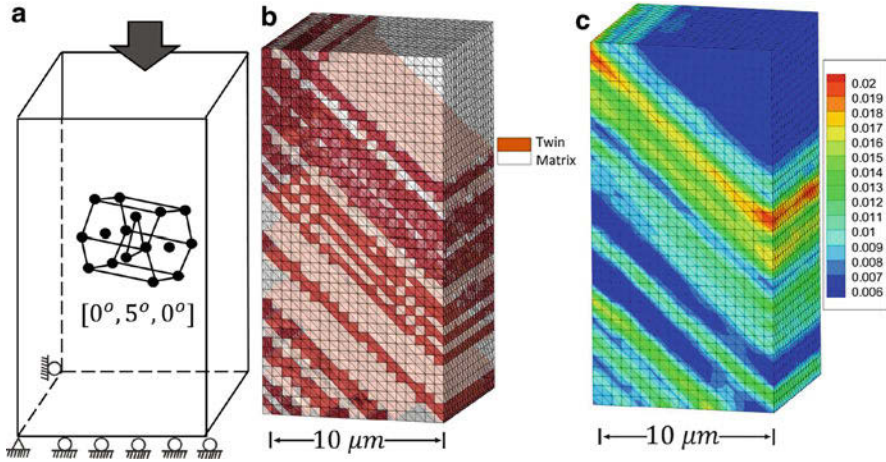
**Fig. 8** (**a**) A pure Mg single crystal computational model subjected to constant strain rate, uniaxial loading, (**b**) simulated twins, and (**c**) Lagrangian strain $E_{yy}$ at 1% strain, using the subcycling-induced CPFE model with a time step $\triangle t = 10\,\text{s}$

1. At time $t$ compute trial displacements and partition coarse and fine sub-domains;
2. Solve the coarse sub-domains with the larger time increment $\triangle t$;
3. Solve the fine sub-domains with scaled time increments $\triangle \tau_i$;
4. From the coarse and fine sub-domain residuals, obtain displacement corrector;
5. Check for equilibrium, update twin domains, and nucleate new twins.

Two factors affect the computational speedup with the subcycling method. They are (i) the ratio of degrees of freedom (DOF) in fine timescale sub-domain to the DOF of entire domain, i.e., $\frac{N^F}{N^{\text{total}}}$, and (ii) the ratio of fine time step $\triangle \tau$ to coarse time step $\triangle t$, i.e., $\frac{\triangle \tau}{\triangle t}$. In Ghosh and Cheng (2018) it is shown that $\frac{\triangle t}{\triangle \tau}$ and $\frac{N^{\text{total}}}{N^F}$ are the key factors in reducing the number of operations. Higher acceleration rates can be achieved by the subcycling method if the deformation is localized in smaller regions and if deformation rates exhibit more heterogeneity. The effectiveness of the twin evolution algorithm with subcycling is tested for single and polycrystalline microstructures.

- **Deformation-Induced Twin Evolution in Single Crystal Mg**: A single crystal simulation of pure Mg is conducted in Cheng and Ghosh (2017) to validate the subcycling-accelerated CPFE model for discrete twin evolution. The microstructure, shown in Fig. 8a, has a dimension of $20 \times 10 \times 10\,\mu\text{m}$ with an Euler angle orientation of $[0°, 5°, 0°]$ in the ZXZ convention. It is discretized into 67,418 TET4 (FP8) elements with 13,021 nodes. The calibrated constitutive parameters for CPFE analysis are provided in Cheng and Ghosh (2017). A uniaxial, constant strain rate of $1 \times 10^{-4}$ is applied in a compressive manner

on the top surface, which causes the formation of $\{10\bar{1}2\}$ tension twins. The 5° tilt in the crystal orientation makes the Schmid factor of twin variant 1 to be highest among all 6 twin variants. Thus only twin variant 1 is formed during simulations.

The evolved discrete twins and Lagrangian strain distribution at 1% strain are shown in Fig. 8b, c, respectively. The $[0°, 5°, 0°]$ orientation causes the $(\bar{1}102)\,[1\bar{1}01]$ twin variant to have the highest Schmid factor. The only exception is that a $(10\bar{1}2)\,[\bar{1}011]$ twin variant (variant 4 of the extension twin systems) occurs at the upper left corner of the model due to the local stress state. The localized strain distribution is caused by easy gliding of twins. Approximately 3.6 times speedup is achieved with subcycling without any loss of accuracy.

- **Twin Evolution in Polycrystalline RVE**: An image-based RVE of the Mg alloy AZ31 containing 620 grains of average grain size of $32\,\mu m$ in a $300 \times 300 \times 300\,\mu m$ box is shown in Fig. 9a. A uniaxial, compressive strain rate of $1 \times 10^{-3}\,s^{-1}$ is applied normal to transverse direction (TD) surface. The figure shows texture with the propagation of twin bands, where the lattice in the twinned region is reorientated by nearly 86°. The volume-averaged stress-strain response from CPFE simulations are compared with experimental results from Beyerlein et al. (2011) in Fig. 9b. The CPFE simulation with subcycling algorithm has a speedup by a factor of 6. In summary, the multi-time-domain subcycling enhanced CPFEM is highly effective for predicting nucleation and propagation of explicit twins in single crystal and polycrystalline microstructures of metals and alloys.

## 5  Adaptive Hierarchical CPFEM with Enhanced Wavelet Basis

Image-based CPFE modeling of polycrystalline materials, e.g., in Roters et al. (2010b), Meissonnier et al. (2001) and Cheng and Ghosh (2017) often requires very high 3D resolution for accurate representation of realistic microstructures. This can incur prohibitively high computational costs in simulating deformation leading to localized phenomena like fatigue failure or deformation twin evolution. A few alternative computational methods have emerged to efficiency-related shortcomings of CPFEM. The elastic-viscoplastic self-consistent models in Lebensohn and Tome (1994) treat grains as embedded inclusions in a homogeneous medium and avoid the need to represent stress heterogeneity inside each grain. The fast Fourier transformation (FFT)-based methods in Moulinec and Suquet (1998) and Lebensohn et al. (2012), and discussed in another chapter of this part, are very efficient, especially for large regular sampling grid simulations with periodicity. The use of FFT methods for discontinuous or high gradient fields can however lead to truncation errors due to the Gibbs phenomenon, propagating from the discontinuity. Regularization methods are being applied to correct this effect, e.g., in Gottlieb et al. (1992). These approaches
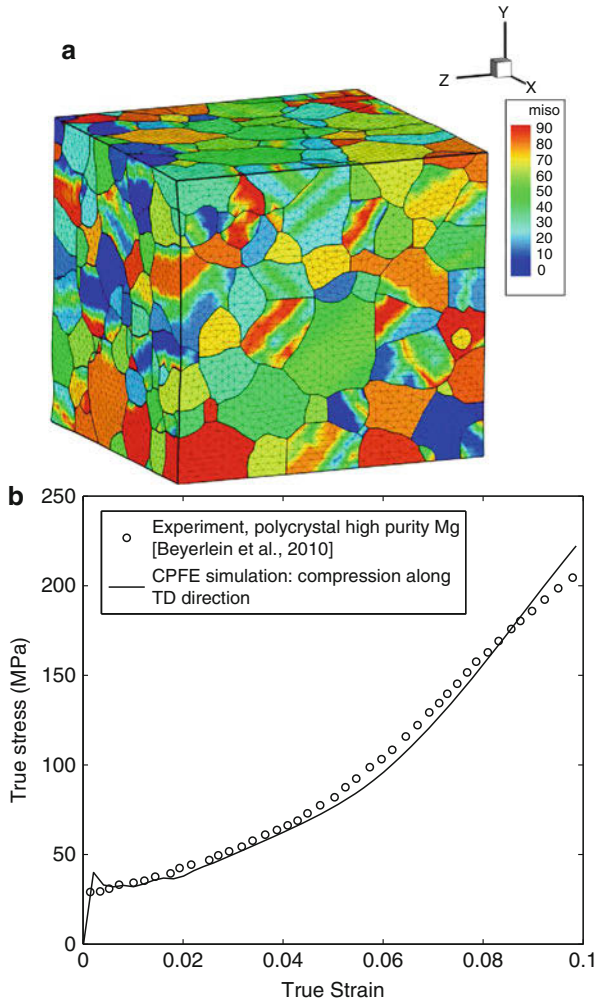
**Fig. 9** (**a**) Texture evolution with deformation in contour plots of the angles between the [0001] lattice axis in each grain and the ND direction (Z-axis) at 2% strain and (**b**) volume-averaged stress-strain response compared with experimental results in Beyerlein et al. (2011)

can also have suboptimal convergence rates due to nonconforming sampling grids that require a large number of degrees of freedom. The present section however seeks an adaptive enrichment method for optimally augment the efficiency of CPFEM while retaining accuracy.

A wavelet-basis enhanced adaptive hierarchical CPFEM has been developed in Azdoud and Ghosh (2017) and Azdoud et al. (2017) to improve computational efficiency and accuracy of FE analyses of polycrystalline microstructures. The method adaptively creates an optimal discretization space conforming to the

solution profile by projecting the solution onto a set of scaling and multi-resolution wavelet basis functions. The multi-resolution property is particularly advantageous for approximating a field using a minimal set of wavelet basis functions. The second generation family of wavelets (Sweldens 1998) is used to generate hierarchical shape functions using the so-called *lifting scheme*. Given the scaling functions at a coarse and fine scale, the lifting scheme defines a set of wavelet functions that complement the coarser set of interpolation functions to uniquely project any function decomposed on a finer set. Complex irregular meshes are easily constructed for these wavelet bases, which make them ideal candidates for enrichment functions in the wavelet-enhanced hierarchical FEM. A summary of the methods and results from Azdoud and Ghosh (2017) and Azdoud et al. (2017) are presented here.

## 5.1 Wavelets for Optimal Enrichment Basis Functions

Wavelet basis functions span the space of square integrable functions $L^2(R)$ through translation and dilation of the scaling function $\varphi(x)$, which satisfies the refinement condition $\varphi(x) = \sum_{k=1}^{N_{\text{filt}}} h_k \varphi(x)(2x - k)$. Parameters $h_k$ and $N_{\text{filt}}$ characterize the wavelet basis and correspond to the components of a low-pass filter. These functions can be used for optimal multi-resolution hierarchical enrichment, conforming to the solution estimate $\tilde{\mathbf{u}}$ profile. Properties that render them ideal for multi-scale enrichment (Azdoud and Ghosh 2017; Azdoud et al. 2017) are:

- *Compact support*: Wavelet functions have compact support. Solutions in wavelet bases do not exhibit spurious instabilities, such as the Gibbs phenomena.
- *Multi-resolution*: Wavelet bases have multi-resolution characteristics that represent the differences between hierarchical scales. Wavelet functions with the Reisz basis property avoid aliasing by ensuring completeness of each scale.
- *Compatibility with FE discretization*: Second-generation wavelet functions (Sweldens 1998) can be constructed from any irregular hierarchical FE mesh.
- *Vanishing moments*: The integral of wavelet functions over any domain is zero. Thus a small coefficient has negligible contribution to the solution.

The lifting scheme creates wavelets for the hierarchical FE shape functions. A *lazy* wavelet $\tilde{\varphi}_\beta^{l-1}$ is first created from hierarchical shape functions, followed by transformation through the lifting scheme by adding vanishing moments to yield:

$$\varphi_\beta^{l-1}(\mathbf{x}) = \tilde{\varphi}_\beta^{l-1}(\mathbf{x}) - \sum_\lambda^R a_\lambda N_\lambda^{l-1}(\mathbf{x}) \tag{27}$$

Here $l$ denotes the scale, $N_\lambda^{l-1}$ is a standard FE shape function at scale $l - 1$ and coefficient $a_\lambda$ is chosen from the condition: $\int_\Omega \varphi_\beta^{l-1}(\mathbf{x}) d\Omega = 0 \quad \forall \beta \in [1, p(l)]$. Each scale of wavelets represents a Reisz basis in $\Omega$. Adding $N_\lambda^{l-1}$

extends the compact support of the wavelet function $\varphi_\beta^{l-1}$ to the whole domain, i.e., $\bigcup_\beta^{p(l)} \text{supp}(\varphi_\beta^{l-1}) = \Omega$. The lifting scheme with $R = 2$ is sufficient for all the above properties.

## 5.2    Adaptive Solution Enhancement with Wavelet Basis Functions

The hierarchical wavelet-enhancement is based on the finite deformation crystal plasticity FE formulation summarized in Sect. 2. The solution in the Newton-Raphson algorithm corresponds to the $i$-th iterative correction $\mathbf{u}$ of the displacement increment $\triangle \mathbf{u}^i$ in a time step between $t$ and $t + \Delta t$. In the proposed algorithm (Azdoud et al. 2017), an adaptive enhancement is made to the first iterate of the solution, i.e., $\Delta \mathbf{u}^{i=1} = \mathbf{u}$. It is premised upon finding an optimal discretization space $V^{h(t)}(\Omega)$ for $\mathbf{u} = \Delta \mathbf{u}^{i=1}$ that will reduce the discretization error to within a prescribed tolerance. Assume that the approximate solution $\mathbf{u}^h$ at time $t$ has been evaluated on the discretized space $V^{h(t)}(\Omega) \subset V(\Omega)$ as:

$$\mathbf{u}^h(\mathbf{x}) = \sum_\alpha^m \mathbf{u}_\alpha^h N_\alpha(\mathbf{x}) + \sum_\beta^{m_{\text{enr}}(t)} \mathbf{w}_\beta^h \varphi_\beta^{(t)}(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega \tag{28}$$

The standard finite element basis $N_\alpha$ corresponds to the approximation of $\mathbf{u}$ in the original coarse FE discretization space at time $t_0$ with $m$ as the number of nodes in this mesh. The adaptive method introduces a set of enrichment functions $\{\varphi\}^{m_{\text{enr}}}$ in the hierarchy, which expand the discretization space $V^h(\Omega)$ to an enriched space $V^{h_{\text{enr}}}(\Omega) \supset V^h(\Omega)$. $m_{\text{enr}}$ corresponds to the number of additional enrichment nodes that are hierarchically added to the initial number $m$. Assume that the set $\{\phi\}^n$ is an arbitrarily large ($n \to \infty$) and sufficient set of multi-scale hierarchical enrichment functions for the coarse discretization space $\{N\}^m$. The functions in the set $\{\phi\}^n$ are the standard $C^0$ hierarchical FEM shape functions obtained by uniform subdivision of the coarse mesh. For the increment $\Delta t \in t \to t + \Delta t$, the adaptive method finds an optimal set $\{\varphi\}^{m_{enr}(t+\Delta t)} \subset \{\phi\}^n$ such that:

$$\|\mathbf{u} - \mathbf{u}^{h_{enr}}\| \le \epsilon \text{ where } \mathbf{u}^{h_{enr}}(\mathbf{x}) = \sum_\alpha^m \mathbf{u}_\alpha^{h_{enr}} N_\alpha(\mathbf{x}) + \sum_\beta^{m_{enr}(t+\Delta t)} \mathbf{w}_\beta^{h_{enr}} \varphi_\beta^{t+\Delta t}(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega \tag{29}$$

An iterative error estimation-solution enrichment algorithm is implemented with iteration steps denoted by $k$. The resulting algorithm for a time step from $t \to t + \Delta t$ has two iterative loops, viz., (i) iterations for the first estimate of $\mathbf{u}$, in which the enrichment functions $\{\varphi\}^{m_{\text{enr}}(t+\Delta t)}$ are sought, and (ii) the Newton-Raphson iterations for the constitutive update. The adaptive iterative scheme for an iteration step $(k + 1)$ to determine $\{\varphi\}^{m_{\text{enr}}(t+\Delta t)}$ is presented in Azdoud et al. (2017).

## 5.3    Results with Adaptive CPFEM for Polycrystals

The effectiveness of the adaptive, wavelet-enhanced hierarchical CPFE model is examined in this example. A polycrystalline microstructure of the *hcp* Ti6242 alloy, containing 208 grains in a computational microstructure of dimensions $124 \times 124 \times 124\,\mu$m is simulated under uniaxial displacement conditions. The $z$-axis misorientation distribution is shown in Fig. 10a. The loading is applied uniformly on the top surface ($z = 124\,\mu$m) with a displacement ramp from $u_z = 0$ to $u_z = 3\,\mu$m, while the surface ($z = 0$) is constrained with $u_x = u_y = u_z = 0$. The transversely
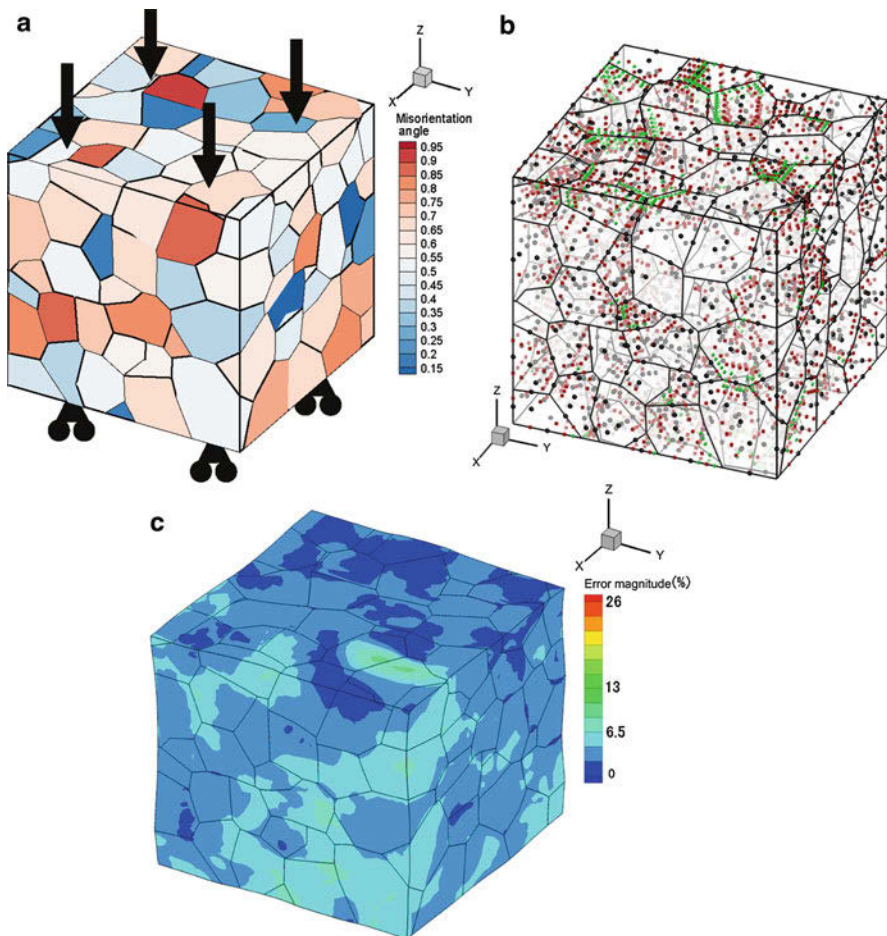


**Fig. 10**  (**a**) A 208 grain polycrystalline microstructure showing $z$-axis misorientation distribution, (**b**) node enrichment positions, where the black spheres, red cubes, and green octahedra denote the enrichment scales 1, 2, and 3, respectively, and (**c**) contour plot of error $e_l(\%)$ for the wavelet adapted model with 42,000 enrichment DoFs
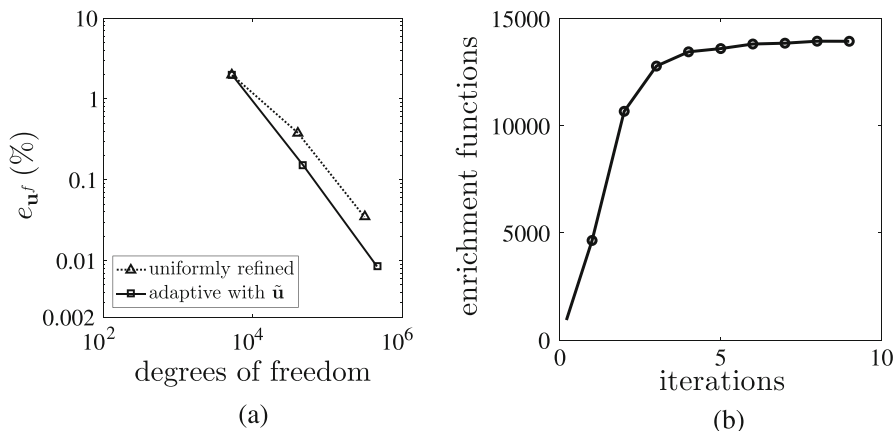
**Fig. 11** (**a**) Log-log plot of the $L_2$ norm of displacement error as a function of DoF and (**b**) evolution of the number of $\tilde{\mathbf{u}}^*$ enrichment functions as a function of the number of iterations

elastic stiffness coefficients and crystal plasticity parameters are given in Azdoud and Ghosh (2017) and Azdoud et al. (2017).

For the elasticity problem only, the wavelet adaptivity admits three scales of enrichment. The coarse-scale solution $\mathbf{u}^h$ is computed on a mesh composed of $N_e = 9485$ tetrahedral elements with $N_n = 1,778$ nodes. The mesh for the fine-scale reference FE solution $\mathbf{u}^f$, as well as the adapted FE model with three scales of enrichment, both have $N_e = 4,856,320$ tetrahedral elements and $N_n = 821,569$ nodes. A total of 14,000 enrichment functions are used, with 5568 functions at scale 1, 6730 functions at scale 2, and 1702 functions at scale 3. In Fig. 10b, nodal positions of the 14,000 enriched function $\{\varphi\}^{m_{\text{enr}}}$ are predominantly in regions of large error near highly misoriented grain boundaries. Figure 10c shows the magnitude of the displacement error $e_l = \sqrt{\frac{(\mathbf{u}^{henr} - \mathbf{u}^f)^T (\mathbf{u}^{henr} - \mathbf{u}^f)}{(\mathbf{u}^f)^T (\mathbf{u}^f)}} \times 100(\%)$ for the adapted solution $\mathbf{u}^{henr}$. The convergence rate of the solution with the adaptive hierarchical FE method is compared to that of a uniformly enriched FE method in Fig. 11a. The average convergence rate for the adaptive method is $\sim \mathcal{O}(N^{-1.22})$ compared to $\sim \mathcal{O}(N^{-0.99})$ for the standard FEM. The evolution of the number of enrichment functions with the number of iterations is depicted in Fig. 11b for a tolerance $\epsilon = 0.002$. Convergence is generally reached in under five iterations. The CPU time for different simulations show a significant gain in efficiency with reduced error for the adaptive method in comparison with a uniformly refined mesh.

For crystal plasticity simulations, convergence rates of the wavelet-enriched adaptive method are compared with that for the uniformly refined hierarchical FEM solutions in Fig. 12a, b. The convergence rates are calculated from displacement and stress errors. The adaptive method converges faster than the uniformly refined hierarchical FEM simulations. For the adaptive method, the convergence rate is $\sim \mathcal{O}(N^{-1.509})$ for the displacement norm and $\sim \mathcal{O}(N^{-0.861})$ for the stress norm,
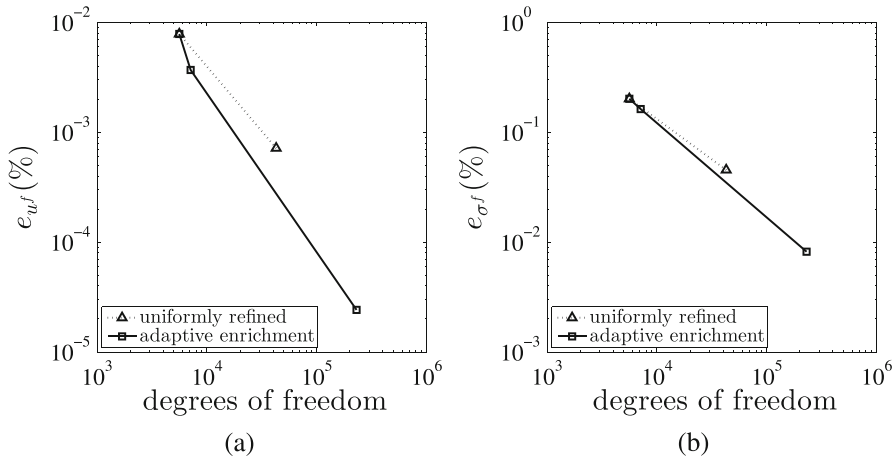
**Fig. 12**  Log-log plot of convergence rates of (**a**) displacement error $e_{\boldsymbol{u}f}$, (**b**) stress error $e_{\boldsymbol{\sigma}f}$ as a function of degrees of freedom at the end of simulation

while those for the uniformly refined model are $\sim\mathcal{O}(N^{-1.179})$ and $\sim\mathcal{O}(N^{-0.733})$, respectively. In conclusion, the proposed adaptive wavelet enrichment method is robust with reduced computational costs while preserving the accuracy of local fields. It is very efficient when a conforming mesh cannot be obtained, such as for heterogeneous materials.

## 6      Conclusions

This chapter emphasizes the need to explore novel methods and algorithms in computational mechanics to facilitate robust and efficient crystal plasticity FE-based modeling of deformation and failure in metals and alloys. Image-based CPFE models incorporate characteristic microstructural features, as well as under-lying physical mechanisms. The claim for such innovative developments is made through three challenging problems that are often bottlenecks in crystal plasticity modeling of extreme mechanisms such as twinning and localization. Remedies to these challenges are developed through methods of element stabilization, multi-time-domain subcycling, and efficiency enhancement through wavelet enhanced hierarchical adaptivity. Many other opportunities exist in the crystal plasticity FE modeling arena. For example, when modeling fatigue failure, it is necessary to conduct simulations for a large number of cycles with a high time resolution to reach local states of crack nucleation and growth. A powerful, multi-resolution wavelet transformation induced multi-time scaling (WATMUS) algorithm has been developed in Chakraborty and Ghosh (2013) and Joseph et al. (2010) for accelerated cyclic simulations. The WATMUS methodology introduces wavelet decomposition of displacements and all associated variables in the finite element formulation to

decouple the response into a monotonic cycle-scale behavior and oscillatory fine timescale behavior. In conclusion, this chapter provides motivation to look beyond available tools and make fundamental advances for effective predictive capabilities. Many of these codes will be hosted at the JHU Software Hub cited in Ghosh (2018).

# References

Abdolvand H, Daymond MR (2013) Multi-scale modeling and experimental study of twin inception and propagation in hexagonal close-packed materials using a crystal plasticity FE approach; part II: local behavior. J Mech Phys Solids 61(3):803–818

Anahid M, Samal MK, Ghosh S (2011) Dwell fatigue crack nucleation model based on crystal plasticity FE simulations of polycrystalline Ti alloys. J Mech Phys Solids 59(10):2157–2176

Asaro RJ, Needleman A (1985) Texture development and strain hardening in rate dependent polycrystals. Acta Mater 33(6):923–953

Azdoud Y, Ghosh S (2017) Adaptive wavelet-enriched hierarchical FE model for polycrystalline microstructures. Comput Methods Appl Mech Eng 321:337–360

Azdoud Y, Cheng J, Ghosh S (2017) Wavelet-enriched adaptive crystal plasticity FE model for polycrystalline microstructures. Comput Methods Appl Mech Eng 321:337–360

Bathe K (2006) Finite element procedures. Prentice Hall/Pearson Education Inc, Upper Saddle River, New Jersey 07458

Beyerlein IJ, McCabe RJ, Tome CN (2011) Effect of microstructure on the nucleation of deformation twins in polycrystalline high-purity magnesium: a multi-scale modeling study. J Mech Phys Solids 59:988–1003

Bridier F, McDowell DL, Villechaise P, Mendez J (2009) Crystal plasticity modeling of slip activity in Ti-6Al-4V under high cycle fatigue loading. Int J Plast 25:1066–1082

Busso EP, Meissonier FT, O'Dowd NP (2000) Gradient-dependent deformation of two-phase single crystals. J Mech Phys Solids 48(11):2333–2361

Chakraborty P, Ghosh S (2013) Accelerating cyclic plasticity simulations using an adaptive wavelet transformation based multitime scaling method. Int J Numer Methods Eng 93(13):1425–1454

Cheng J, Ghosh S (2015) A crystal plasticity FE model for deformation with twin nucleation in magnesium alloys. Int J Plast 67:148–170

Cheng J, Ghosh S (2017) Crystal plasticity FE modeling of discrete twin evolution in polycrystalline Magnesium. J Mech Phys Solids 99:512–538

Cheng J, Shahba A, Ghosh S (2016) Stabilized tetrahedral elements for crystal plasticity FE analysis overcoming volumetric locking. Comput Mech 57(5):733–753

Cheng J, Shen J, Mishra RK, Ghosh S (2018) Discrete twin evolution in Mg alloys using a novel crystal plasticity finite element model. Acta Mater 149:142–153

de Souza Neto EA, Andrade Pires FM, Owen DRJ (2005) F-bar-based linear triangles and tetrahedra for finite strain analysis of nearly incompressible solids. Part I: formulation and benchmarking. Int J Numer Methods Eng 62:353–383

Deka D, Joseph DS, Ghosh S, Mills MJ (2006) Crystal plasticity modeling of deformation and creep in polycrystalline Ti-6242. Metall Trans A 37(5):1371–1388

Dohrmann CR, Heinstein MW, Jung J, Key SW, Witkowski WR (2000) Node-based uniform strain elements for 3-node triangular and 4-node TET meshes. Int J Numer Methods Eng 47: 1549–1568

Dunne FPE, Kiwanuka R, Wilkinson AJ (2012) Crystal plasticity analysis of micro-deformation, lattice rotation and geometrically necessary dislocation density. Proc R Soc Lond A 468: 2509–2531

Gee MW, Dohrmann CR, Key SW, Wall WA (2009) A uniform nodal strain tetrahedron with isochoric stabilization. Int J Numer Methods Eng 78:429–443

Ghosh S (2018) JH-SofHub: Johns Hopkins University Software Hub. https://jhsofthub.wse.jhu.edu/about-2/

Ghosh S, Cheng J (2018) Adaptive multi-time-domain subcycling for crystal plasticity FE modeling of discrete twin evolution. Comput Mech 61(1):33–54

Gottlieb D, Shu CW, Solomonoff A, Vandeven H (1992) On the Gibbs phenomenon I: recovering exponential accuracy from the Fourier partial sum of a nonperiodic analytic function. J Comput Appl Math 43(1):81–98

Hasija V, Ghosh S, Mills MJ, Joseph DS (2003) Modeling deformation and creep in Ti-6Al alloys with experimental validation. Acta Mater 51:4533–4549

Izadbakhsh A, Inal K, Mishra RK, Niewczas M (2011) New crystal plasticity constitutive model for large strain deformation in single crystals of magnesium. Model Simul Mater Sci Eng 50: 2185–2202

Joseph DS, Chakraborty P, Ghosh S (2010) Wavelet transformation based multi-time scaling method for crystal plasticity FE simulations under cyclic loading. Comput Methods Appl Mech Eng 199(33):2177–2194

Kalidindi SR, Schoenfeld SE (2000) On the prediction of yield surfaces by the crystal plasticity models for FCC polycrystals. Mater Sci Eng A 293:120–129

Keshavarz S, Ghosh S (2013) Multi-scale crystal plasticity FEM approach to modeling Nickel based superalloys. Acta Mater 61:6549–6561

Keshavarz S, Ghosh S (2015) Hierarchical crystal plasticity FE model for Nickel-based superalloys: sub-grain microstructures to polycrystalline aggregates. Int J Solids Struct 55:17–31

Knezevic M, Daymond MR, Beyerlein IJ (2016) Modeling discrete twin lamellae in a microstructural framework. Scripta Mater 121:84–88

Kocks UF, Argon AS, Ashby MF (1975) Thermodynamics and kinetics of slip. Prog Mater Sci 19:141–145

Lebensohn RA, Tome CN (1994) A self-consistent viscoplastic model: prediction of rolling textures of anisotropic polycrystals. Mater Sci Eng A 175:71–82

Lebensohn RA, Kanjarla AK, Eisenlohr P (2012) An elasto-viscoplastic formulation based on fast Fourier transforms for the prediction of micromechanical fields in polycrystalline materials. Int J Plast 32–33:59–69

Matous K, Maniatty AM (2004) FE formulation for modeling large deformations in elasto-viscoplastic polycrystals. Int J Numer Methods Eng 60:2313–2333

Meissonnier FT, Busso EP, O'Dowd NP (2001) FE implementation of a generalised non-local rate-dependent crystallographic formulation for finite strains. Int J Plast 17: 601–640

Moulinec H, Suquet P (1998) A numerical method for computing the overall response of nonlinear composites with complex microstructure. Comput Methods Appl Mech Eng 157(1–2):69–94

Nguyen-Thoi T, Liu GR, Lam KY, Zhang GY (2009) A face-based smoothed FE method (FS-FEM) for 3D linear and non-linear solid mechanics problems using 4-node tetrahedral elements. Int J Numer Methods Eng 78:324–353

Ozturk D, Pilchak AL, Ghosh S (2017) Experimentally validated dwell fatigue crack nucleation model for alphaTi alloys. Scripta Mater 127:15–18

Pierce D, Asaro RJ, Needleman A (1983) Material rate-dependence and localized deformation in crystalline solids. Acta Metall 31(12):1951–1976

Roters F, Eisenlohr P, Bieler TR (2010a) Crystal plasticity FE methods in materials science and engineering. Wiley-VCH Verlag GmbH, Weinheim

Roters F, Eisenlohr P, Hantcherli L, Tjahjantoa DD, Bieler TR, Raabe D (2010b) Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity FE modeling: theory, experiments, applications. Acta Mater 58(4):1152–1211

Shahba A, Ghosh S (2016) Crystal plasticity FE modeling of Ti alloys for a range of strain-rates. Part I: a unified constitutive model and flow rule. Int J Plast 87:48–68

Sinha V, Mills MJ, Williams JC, Spowart JE (2006) Observations on the faceted initiation site in the dwell-fatigue tested Ti-6242 alloy: crystallographic orientation and size effects. Metall Mater Trans A 37(5):1507–1518

Staroselsky A, Anand L (2003) A constitutive model for hcp materials deforming by slip and twinning: application to magnesium alloy AZ31B. Int J Plast 19:1843–1864

Sweldens W (1998) The lifting scheme: a construction of second generation wavelets. SIAM J Math Anal 29(2):511–546

Thomas J, Groeber M, Ghosh S (2012) Image-based crystal plasticity FE framework for microstructure dependent properties of Ti-6Al-4V alloys. Mater Sci Eng A 553:164–175

Venkataramani G, Ghosh S, Mills MJ (2007) A size dependent crystal plasticity FE model for creep and load-shedding in polycrystalline Titanium alloys. Acta Mater 55:3971–3986

Zhang J, Joshi SP (2012) Phenomenological crystal plasticity modeling and detailed microme-chanical investigations of pure magnesium. J Mech Phys Solids 60:945–972

# Part XI
# Materials Informatics

# Materials Informatics: Overview

# 80

Nicola Marzari

## Contents

**Abstract**

Recent years have seen the explosion of a field of research now commonly termed "materials informatics." Though the roots of this research can be traced back to the early 1970s, and mostly driven by the chemistry community, the last 20 years have seen the pioneering and then systematic application of deterministic or stochastic methods based on large data collections to design or discover novel materials, where the materials data themselves are the outcome of calculations, or exploit these in synergy with experimental databases. This chapter aims to provide an overview of some of the most successful and exciting efforts worldwide in this area, with a focus on materials science and condensed-matter physics, but also including notable contributions in chemistry and molecular science. Contributions can be broadly assigned to two different areas: efforts dedicated to producing and/or storing curated or raw computational or experimental data and associated materials properties, and machine-learning

N. Marzari (✉)
Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: nicola.marzari@epfl.ch

efforts dedicated to leveraging those data to bypass the need of expensive, and typically first-principles, calculations. Both hold great promise in our quest to understand, predict, and design the properties of novel materials.

# 1    Introduction

Adamson and Bush (1973, 1974) noted how "some chemical information systems hold property data as well as structure diagrams in machine-readable forms," and suggested that "if structure-property relationships could be investigated easily within chemical information systems, then the usefulness of the systems would be considerably increased." In 1980, Marvin Cohen and collaborators (▶ Chap. 3, "Modeling Solids and Its Impact on Science and Technology") were able to show that computational condensed-matter physics and materials science had reached the capability to predict materials properties entirely from first-principles, i.e., without any experimental input. The roots of what is happening today are all here: calculations can predict materials data, and this materials data can be used to predict materials properties, which is almost a revolution. For all good revolutions, a cadre of revolutionaries is needed. I will list three in the vanguard that I found inspiring. Franceschetti and Zunger (1999) introduced the materials community to the concept of inverse design – by capturing a structure-property relationship in a model (something that had been done for quite some time, e.g., using cluster, Ising, or Heisenberg expansions to predict the energetics of alloys, or the phase transitions of magnetic materials), it becomes possible to focus on the inverse design problem – rather than predicting the properties of a structure, search for the structure that could deliver a target property. Jóhannesson et al. (2002) showed how first-principles calculations can be executed reliably in a high-throughput manner, allowing to explore and identify novel and promising metal alloys. Curtarolo et al. (2003) showed how one can leverage the calculation of large sets of first-principles materials data to extract heuristic rules that could then be applied to, e.g., crystal structure prediction. Therefore, inverse design, high-throughput, and data mining all make their appearance in the materials world.

The unsung heroes of this revolution are, of course, the computer codes that make it all possible. The calculation of complex materials properties from first-principles required, and requires, efforts that span decades and involved dedicated individuals and groups, albeit rewarded by the enormous impact that these tools have on the community. (*Nature* identified, in 2014, 12 papers dedicated to density-functional theory among the 100 most-cited papers ever in the field of science, engineering, and medicine/life sciences (Van Noorden et al. 2014).) Without these computer codes, it would not have been possible to translate the pioneering ideas mentioned earlier into an industrial revolution of computing farms incessantly and inexorably extending and refining databases of calculated materials properties, and of intelligent computer agents condensing that knowledge in parametric and nonparametric forms to predict even more, or even less expensively, the properties. Given the exponential growth of throughput capacity for the past 30 years, doubling every 14 months, the future

seems exceedingly predictable – less hype than reality, materials informatics is here to stay.

---

## 2    Contributed Chapters to "Materials Informatics"

Contributions presented in this chapter offer broad vistas for the field and highlight some of the most relevant efforts in producing, collecting, and curating computational data; in collecting and curating experimental data; and in leveraging either resource with machine learning.

Some of the early efforts in calculating and providing large amounts of computational data for materials, alluded to in the Introduction, are represented by the Materials Project (https://materialsproject.org/), based at LLNL; by the AFLOW consortium (http://aflowlib.org/), based at Duke University; by the Open Quantum Materials Database (OQMD, http://oqmd.org/), based at Northwestern University; by the Computational Materials Repository (CMR, https://cmr.fysik.dtu.dk/), based at DTU; and by the Materials Cloud (https://www.materialscloud.org/), based at EPFL and CSCS. All these are discussed in this section, and several are aggregated in the NoMaD repository (https://nomad-repository.eu/), based at the Fritz Haber Institute (for an overview, including chemical and molecular repositories, see the contribution in ▶ Chap. 88, "Machine Learning and Big-Data in Computational Chemistry"). All these repositories support an open-access model to the data, although the complexity of bulk retrieval may vary. Notable, in this respect, has been the creation of the OptiMaDe consortium (http://www.optimade.org), collecting all the above players, with the stated goal of making all the materials databases interoperational by developing a common REST API.

Of course, the field of materials repositories predates computational efforts, and a long-term visionary effort has been in place for decades to curate experimental structures and properties into the Pauling File (http://paulingfile.com/), together with its recent web frontend (https://mpds.io/). The Pauling File is also notable for having collected through the years an extensive and standardized dictionary of materials properties (http://paulingfile.com/index.php?p=physical%20properties) that could be very helpful for the ontologies being developed nowadays, spearheaded by the European Materials Modelling Council (https://emmc.info/) and the RDA Task Group for Semantic Assets for Materials Science. The Inorganic Crystal Structure Database (ICSD, http://www2.fiz-karlsruhe.de/icsd_home.html) provides also an extensive database of materials' structures, and has been used extensively as a starting point for the calculations in the Materials Project, AFLOW, and OQMD. At variance with these two previous efforts, where data are licensed, the Crystallography Open Database (COD, http://www.crystallography.net/cod/) is an open-access collection of crystal structures of organic, inorganic, and metal-organic compounds and minerals, freely accessible.

The effort to drive thousands of calculations (or even millions, as is the case of AFLOW) has had a second major benefit – i.e., it has driven the development of the infrastructure needed to execute, run, and monitor calculations in

fully automatic ways – this is implemented with different levels of complexity in ASE (https://wiki.fysik.dtu.dk/ase/), AiiDA (http://www.aiida.net), Atomate (https://atomate.org), AFLOWπ (http://aflowlib.org/src/aflowpi/), or Abipy (https://github.com/abinit/abipy), to name a few starting with A. Such capabilities will make it increasingly straightforward for other computational modeling groups or experimental teams to rapidly benefit from the usage of complex electronic-structure codes, with modeling services even available freely when open-source codes are used (e.g., the model of the Work section of the Materials Cloud).

Python has also emerged as the de facto standard for all these projects and has been adopted by the community following the early efforts of ASE and PyMatGen (http://pymatgen.org/) that have now become power tools for the analysis of the materials data produced by the simulations, leading to the automatic construction of complex materials relationships, such as phase or Pourbaix diagrams.

Several efforts are showcased here, with an extensive description of the computational infrastructure being developed in the Materials Project (▶ Chap. 81, "The Materials Project: Accelerating Materials Design Through Theory-Driven Data and Tools"), in the AFLOW consortium (▶ Chap. 82, "The AFLOW Fleet for Materials Discovery"), and in the Swiss and European Centres MARVEL (http://nccr-marvel.ch) and MaX (http://www.max-centre.eu/) through the AiiDA materials' informatics framework and the Materials Cloud dissemination platform (▶ Chap. 83, "Open-Science Platform for Computational Materials Science: AiiDA and the Materials Cloud").

The curation of experimental datasets is discussed in the two contribution from the Pauling File (▶ Chap. 84, "The PAULING FILE Project and Materials Platform for Data Science: From Big Data Toward Materials Genome") and COD (▶ Chap. 85, "Crystallography Open Database (COD)"), that often, together with ICSD and the chemistry databases, provide the starting point, the verification step, and the variety that is needed for the discovery or prediction projects.

Last, and most excitedly, the wealth of information becoming available makes it possible to data mine the current resources, to uncover unexpected correlations, or to provide novel suggestions for optimal materials, but especially to short-circuit the entire expensive electronic-structure framework by machine learning the simple or complex properties – from potential energy surfaces to NMR chemical shifts – that would otherwise be obtained with calculations that are several orders of magnitude more expensive. It can be envisioned that even the most expensive quantum chemistry and wave function methods in electronic-structure – providing the highest degree of reliability and predictive accuracy – could powerfully provide the baseline for machine learning, thus reaching what has been up to now the impossible dream of modeling complex properties or materials in complex environments with full quantum-mechanical accuracy.

The field is here extensively represented by contributions spanning the world of materials and of molecules, and bridging them. In ▶ Chap. 86, "Quantum Machine Learning in Chemistry and Materials," the fundamental mathematical formulations of quantum machine learning (i.e., of models which can be used in the context of quantum-mechanical training and testing data) are discussed, highlighting also

how up to now the primary focus has been on ground-state or local minimum properties, rather than more challenging excited states or properties that depend in an ill-conditioned manner on atomic coordinates. In ▶ Chap. 87, "Machine Learning of Atomic-Scale Properties Based on Physical Principles," a mathematical framework is presented, based on the concept of the atomic density, for building descriptors of atomic environments that preserve the geometric symmetries and chemically sensible limits. Coupled with kernel regression, this allows the fitting of complex models of physical properties on the atomic scale, both scalars such as interatomic potentials (force fields) and tensors such as multipole moments and quantum mechanical operators. In ▶ Chap. 88, "Machine Learning and Big-Data in Computational Chemistry", a broad perspective on current efforts in machine learning, and materials screening and discovery, is provided. Notably, it also underscores how computational modeling is becoming increasingly a high-velocity and high-volume field, but where variety is somewhat lacking, leading to the need to address heterogeneous, unstructured data to recover variety.

## 3 Final Concluding Remarks

Some considerations on data production, collection, and exploitation are then warranted. One might go as far as highlighting first a core difference in the world of data, and identify three broad data categories:

- Social data (harvested, most often in uncontrolled conditions)
- Experimental data (harvested, in largely controlled conditions)
- Computational data (generated, in controlled conditions)

The fact that in computational science the data are calculated, under controlled conditions, reminds us of a core difference with respect to other fields – i.e., putting squarely at the center of this effort the codes and workflows of different codes that are able to produce automatically, and on-demand, the desired quantity. And each of these quantities has its own pedigree – how expensive it is to calculate, how much intermediate data are needed; the balance between the cost of calculating and the cost of storing being property and material specific. But, of course, one needs data to begin with, and so right now there is a healthy and biodiverse ecosystem of curated, raw, and aggregated data that explore all the different models – all probably needed, and all useful in different ways. If one wanted a simplification, one could say that the most pressing needs in computational sciences are those of having the following:

1. Highly curated data (fully verified, with standardized protocols)
2. Data on demand (the capability of calculating new data on an automatic request)

The interplay of theory, algorithms, and workflows that are able to deliver these goals are able to produce the vast trove of simple or complex materials'

properties that can then be data mined or machine learned – these not only are great accelerators, but promise to direct searches in directions that would not have been intuitive to soft-matter intelligence.

These considerations bring us to underscore the power that machine learning will have in materials research. Here, three easy predictions are made. First, the emergence of parametric and non-parametric models, such as neural network and Gaussian approximation potentials, has delivered the remarkable recognition that it is possible to reproduce with great accuracy and at costs that are reduced by 4– 5 orders of magnitude of the potential energy surfaces of materials (Behler and Parrinello 2007; Bartók et al. 2010; Li et al. 2015). This has, and will have, a revolutionary impact in those fields where deterministic or stochastic sampling of materials relies on length and time scales a few orders of magnitude larger than current capabilities, and with an accuracy that only first-principles methods can deliver. This is very much the domain of metals' plasticity, or nucleation and growth – both fields underpin all our efforts in understanding and engineering synthesis and manufacturing of advanced materials – from vapor-liquid-solid growth of nanostructures to fast precipitation in additive manufacturing. Second, the capability to machine-learn properties that are very expensive to calculate – from thermodynamics with chemical accuracy to magnetic spectroscopies or transport properties – will provide a great accelerator in the direct, high-throughput discovery or the inverse design of materials. Last, harvesting and correlations can take place in the social domain – natural language processing aims to identify properties or even chatter in the editorial or social world – and in the experimental domain, so one can easily envision an ever-growing effort to leverage heterogeneous information from these domains, and combine with homogenous (but systematically biased, in mysterious ways) computed data to aim for that elusive, but most desirable, breakthrough of a novel material able to contribute, hopefully, to the growth and well-being of our society.

# References

Adamson GW, Bush JA (1973) A method for the automatic classification of chemical structures. Inform Stor Retr 9:561

Adamson GW, Bush JA (1974) Method for relating the structure and properties of chemical compounds. Nature 248:406

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98:146401

Curtarolo S, Morgan D, Persson K, Rodgers J, Ceder G (2003) Predicting crystal structures with data Mining of Quantum Calculations. Phys Rev Lett 91:135503

Franceschetti A, Zunger A (1999) The inverse band structure problem: find the atomic configuration with given electronic properties. Nature 402:60

Jóhannesson GH, Bligaard T, Ruban AV, Skriver HL, Jacobsen KW, Nørskov JK (2002) Combined electronic structure and evolutionary search approach to materials design. Phys Rev Lett 88:255506

Li Z, Kermode JR, De Vita A (2015) Molecular dynamics with on-the-Fly machine learning of quantum-mechanical forces. Phys Rev Lett 114:096405

Van Noorden R, Maher B, Nuzzo R (2014) The top 100 papers. Nature 514:550

# The Materials Project: Accelerating Materials Design Through Theory-Driven Data and Tools

# 81

Anubhav Jain, Joseph Montoya, Shyam Dwaraknath, Nils E. R. Zimmermann, John Dagdelen, Matthew Horton, Patrick Huck, Donny Winston, Shreyas Cholia, Shyue Ping Ong, and Kristin Persson

## Contents

A. Jain (✉) · J. Montoya · S. Dwaraknath · N. E. R. Zimmermann · M. Horton · P. Huck · D. Winston · S. Cholia
Lawrence Berkeley National Laboratory, Berkeley, CA, USA
e-mail: ajain@lbl.gov; montoyjh@lbl.gov; shyamd@lbl.gov; nerz@lbl.gov; mkhorton@lbl.gov; phuck@lbl.gov; dwinston@lbl.gov; scholia@lbl.gov

J. Dagdelen · S. P. Ong
University of California, Berkeley, CA, USA
e-mail: jdagdelen@berkeley.edu; ongsp@eng.ucsd.edu

K. Persson
Lawrence Berkeley National Laboratory, Berkeley, CA, USA

University of California, Berkeley, CA, USA
e-mail: kapersson@lbl.gov

**Abstract**

The Materials Project (MP) is a community resource for theory-based data, web-based materials analysis tools, and software for performing and analyzing calculations. The MP database includes a variety of computed properties such as crystal structure, energy, electronic band structure, and elastic tensors for tens of thousands of inorganic compounds. At the time of writing, over 40,000 users have registered for the MP database. These users interact with this data either through the MP web site (https://www.materialsproject.org) or through a REpresentational State Transfer (REST) application programming interface (API). MP also develops or contributes to several open-source software libraries to help set up, automate, analyze, and extract insight from calculation results. Furthermore, MP is developing tools to help researchers share their data (both computational and experimental) through its platform. The ultimate goal of these efforts is to accelerate materials design and education by providing new data and software tools to the research community. In this chapter, we review the history, theoretical methods, impact (including user-led research studies), and future goals for the Materials Project.

# 1    History and Overview of the Materials Project

Materials scientists and engineers have always depended on materials property data to inform, guide, and explain research and development. Traditionally, such data originated almost solely from experimental studies. In the past 10–15 years, it has become possible to rapidly generate reliable materials data using scalable computer simulations of the fundamental equations of physics such as the Schrödinger equation. This paradigm shift was induced by a combination of theoretical advances, most notably the development of density functional theory (DFT), algorithmic improvements, and low-cost computing.

The Materials Project (MP, or "The Project") was founded in 2011 as a collaborative effort to leverage ongoing advances in theory and computing to accelerate the research and design of new materials. The Project rests on a comprehensive database of predicted properties of materials that is the result of executing millions of DFT simulations on supercomputing resources. At the time of writing, this database includes $>69,000$ inorganic materials with crystal structures and total energies, $>57,000$ materials with electronic band structures, $>48,000$ with electronic transport properties (Fig. 1) (Ricci et al. 2017), $>30,000$ with XANES k-edge spectra (Dozier et al. 2017), $>15,000$ with conversion battery properties, $>6000$ with elastic tensors (de Jong et al. 2015a), $>3,000$ with intercalation battery
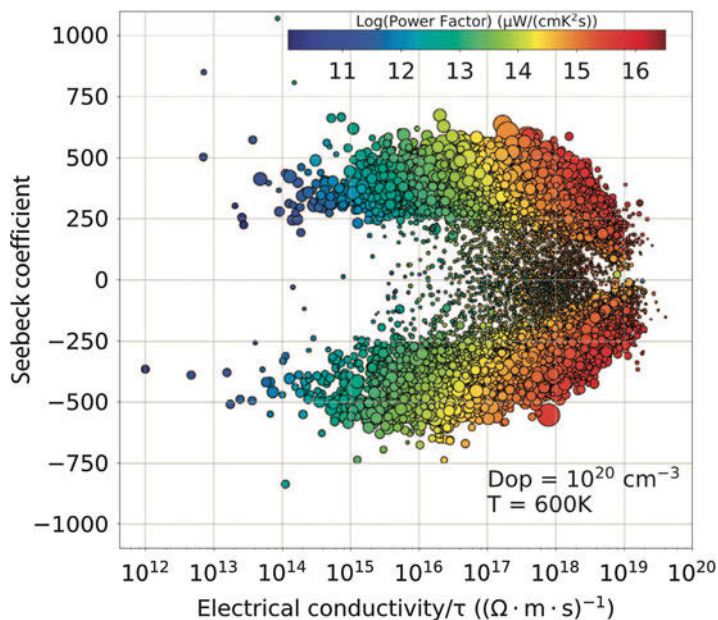
**Fig. 1** Example of a large electronic transport data set in MP generated through computations. Each point represents one compound, with Seebeck coefficient versus electron conductivity (divided by $\tau$) plotted. The color represents the thermoelectric power factor ($S^2\sigma$), and the point size is proportional to the bandgap (Ricci et al. 2017). This data set is available through the MPContribs platform (see Sect. 6.2) at: https://materialsproject.org/mpcontribs/boltztrap

properties, >1,000 with piezoelectric tensors (de Jong et al. 2015b), >1,000 with dielectric tensors (Petousis et al. 2017), and > 1000 elemental surface energies (Tran et al. 2016). This database is continually expanding with more materials and more properties (see Fig. 2 for an example of properties listed in the current iteration).

The Project launched its publicly accessible web site in October 2011 and has since grown into a multi-institution collaboration as part of the US Department of Energy Office of Basic Energy Sciences (BES). The web site provides access to the database as well as applications (or "apps") that combine and visually present the data for specific analyses such as phase diagram generation or battery electrode evaluation. The MP web site hosts more than 40,000 registered users worldwide consisting of a diverse set of researchers and students from academia, industry, and educational institutions (Figs. 3 and 4). The diversity of the audience base highlights the usefulness of a theory-based materials database across the spectrum of education, research, and development activities.

Apart from the core data and web site, MP helps develop and maintain a set of open-source software libraries for setting up, executing, analyzing, and deriving insights from calculations. These libraries, which include pymatgen (Ong et al. 2013), custodian, FireWorks (Jain et al. 2015), and atomate (Mathew et al. 2017),

**Fig. 2** An example of a "materials detail" page for BaTiO$_3$ on the Materials Project web site. The information available includes crystal structure parameters, thermodynamic properties, electronic band structure and density of states, a simulated x-ray diffraction pattern, simulated x-ray absorption spectra, a substrate matcher, an elasticity tensor, a piezoelectric tensor, a table and links pertaining to calculation methodology details, and metadata regarding the crystal structure source
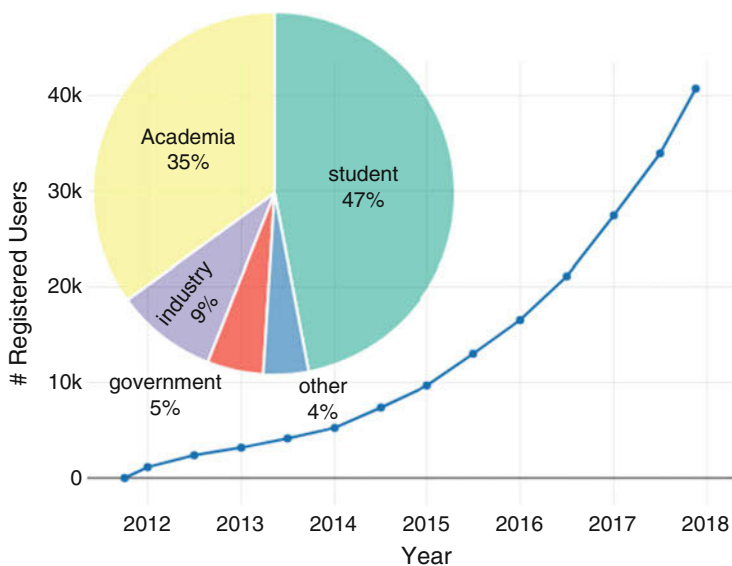
**Fig. 3** Total number of registered users since release of the MP web site and fraction of users belonging to various institution types
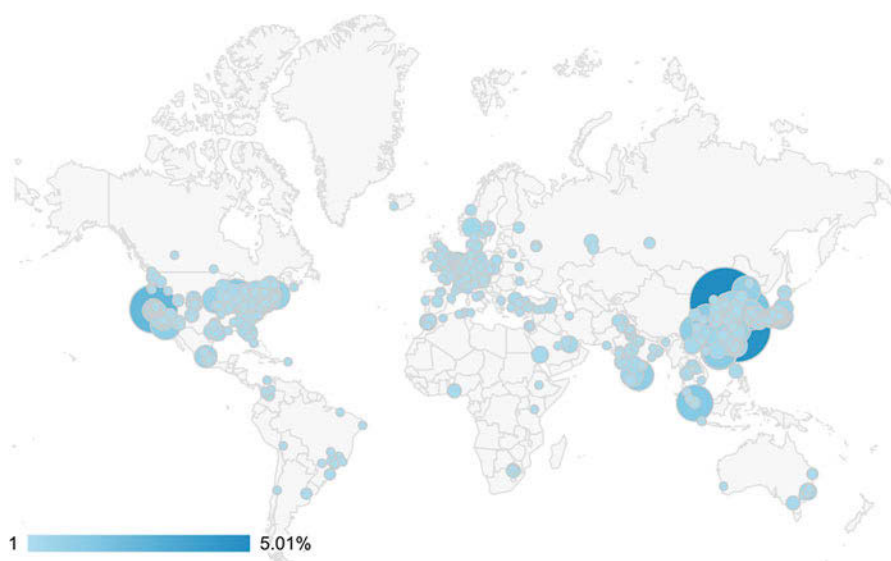


**Fig. 4** Amount of Materials Project user sessions by city for the month of October 2017. Sessions originated in 112 countries, 36 of which totaled >100 sessions

have been used by hundreds of researchers worldwide. The newest additions to MP allow users to suggest compounds for computation as well as contribute their own data (theoretical or experimental) to the database. Furthermore, MP hosts educational workshops focused on its online and programmatic infrastructure, and the MP web site has become an integral teaching tool in several materials science courses.

As the state of the art in theory and computing are bound to change, the specifics of MP's data, scope, capabilities, and infrastructure will no doubt change as well. Nevertheless, this chapter summarizes the current state of the Materials Project.

## 2 Underlying Theoretical Formalism and Development of Materials Design "Apps"

### 2.1 Theoretical Methods

The Materials Project's core data set consists of results obtained from density functional theory (DFT) calculations on a library of inorganic compounds. DFT is well suited for creating a database of materials properties because it has fewer parameters that require tuning for different materials systems and because the computational cost for small- to medium-sized (approximately 300 atoms or less) systems is manageable. DFT methods have become standardized to a large extent such that various software implementations with slightly different parameters (e.g., pseudopotentials) produce very similar results (Lejaeghere et al. 2016).

Nevertheless, selecting a robust set of parameters for high-throughput computations is still not trivial. It is important to emphasize there is currently no perfect DFT functional as they are all approximations to the complete set of physics that define materials phenomena. For example, strongly correlated systems remain challenging. It is typically possible to treat even complex systems with specialized methods in single studies. However, when constructing a large database with many compounds, such specialized treatment is difficult to achieve practically and would also lead to inconsistent and often incompatible results between various compounds. Additionally, one must more carefully balance computational costs with expected information gain. Whereas a single study may not be noticeably impacted if its calculations are over-converged numerically and use 50% more computing power than necessary, such a situation would severely slow down a high-throughput database project such as MP that consumes tens of millions of CPU hours of computing per year. Thus, MP must make practical compromises that try to maintain the accuracy of a specialized, precise calculation while being completely automatic and computationally efficient and maintaining clarity and consistency of procedure with other calculations.

One of the approaches used by the Materials Project to achieve this balance is to split materials into two classes and apply a different DFT functional to model each

class. The first class of compounds are transition metal oxides and sulfides. Standard DFT functionals such as the local density approximation (LDA) (Kohn and Sham 1965) and the generalized gradient approximation (GGA) (Perdew et al. 1996) are not accurate for these compounds due to more pronounced self-interaction error as well as errors in orbital occupation from lack of derivative discontinuity (Zhou et al. 2004; Cococcioni and de Gironcoli 2005). One computationally efficient way to treat these compounds is with the GGA+$U$ framework, in which a Hubbard-like correction is applied to localized $d$ orbitals. The specific $U$ corrections are fitted to formation energy data as described previously (Wang et al. 2006). It is important to note that these same $U$ values may not be optimal for accurately representing other properties such as the electronic band structure. The second class of compounds encompasses all other systems and is treated with the standard GGA-PBE functional (Perdew et al. 1996).

By allowing different compounds to be treated with two different functionals, it is possible to enhance accuracy of the resulting database compared to using only a single functional such as GGA for the entire database. However, one must then additionally design a scheme to mix results (e.g., total energies) obtained from different methods since these results are not directly compatible. In the Materials Project, these adjustments between results from different functionals are made by benchmarking to experimental formation enthalpy data (Jain et al. 2011b). Figure 5 depicts the effects of one instance of this by presenting Fe-P-O phase diagrams using the GGA only, GGA+$U$ only, and mixed GGA and GGA+$U$ total energies. Only the version of the diagram that uses two different functionals (with the mixing adjustment applied) reproduces all known stable phases in this system.

Another practical measure taken by the Materials Project pertains to molecular systems. Although molecular systems and solids can be modeled within the same density functional theory framework (e.g., PBE-GGA with plane-wave basis sets), computed reaction energies that include both molecules and solids typically exhibit high errors because self-interaction errors differ significantly between systems characterized by local (e.g., molecules or highly correlated systems) and nonlocal (e.g., metals) electrons (Grindy et al. 2013; Perdew et al. 1998). Similarly, intermolecular interactions present in gases, 2D materials, and liquids that are not well described by pure GGA functionals present further challenges for constructing a comprehensive thermodynamic framework derived from DFT that avoids such systematic errors.

Rather than calculating the liquid/gas energies directly, MP adjusts the energies of several elements that are liquid or gaseous at room temperature based on experimental reaction enthalpies such as the oxidation of metals (Wang et al. 2006). All of the following compounds have adjusted energies to better reproduce reaction energies with solid phases: $O_2$, $N_2$, $Cl_2$, $F_2$, and $H_2$.

Finally, we mention that MP also adjusts certain numerical parameters based on the type of compound. For example, MP uses a denser k-point mesh when calculating metals (as determined from an initial, loose k-point mesh calculation) versus semiconductors and insulators. In addition, the numerical tolerances used by the Materials Project have been growing more precise over time. The parameters
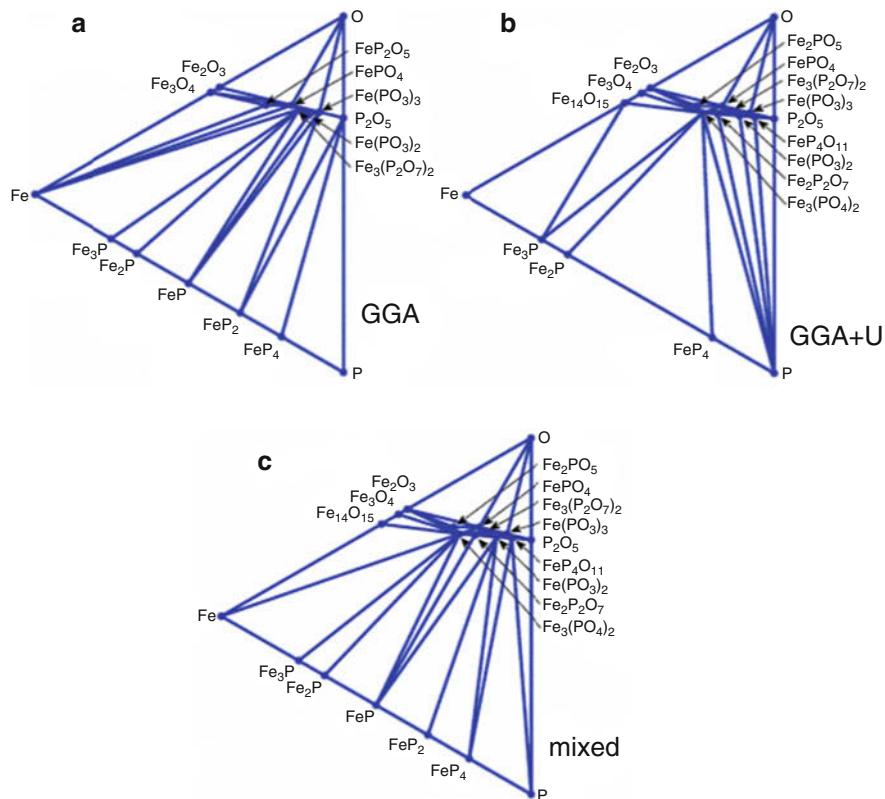
**Fig. 5** Fe-P-O ternary phase diagrams built using total energy calculations from (**a**) only GGA, (**b**) only GGA+U, and (**c**) mixing GGA and GGA+U functionals. Only the mixed phase diagram reproduces all known phases as stable on the phase diagram. (Reprinted figure with permission from Jain et al. (2011b). Copyright 2011 by the American Physical Society)

used for each calculation are available via the Materials Project web site, and the most current description of parameter settings is provided at https://www.materialsproject.org/docs/calculations.

## 2.2 "Apps" for Data Exploration

Much of the value of the MP data set comes from secondary analyses that are performed on top of the raw data. These secondary analyses often combine multiple data points and can take the form of common diagrams used in materials science (e.g., phase stability diagrams or Pourbaix diagrams), application-specific materials design tools (e.g., evaluating MP compounds as battery electrodes), or simply as additional information (e.g., reporting potential substrates that might form coherent

lattices with a target material). Such tools are vital for helping users extract as much value as possible from the data sets.

The Materials Project develops the methodologies to perform many such secondary analyses and releases them both as open-source software implementations (through the pymatgen (Ong et al. 2013) package) and as web applications ("apps"). Apps provide a visual, user-friendly interface to these powerful and often complex analysis routines. In the following example, we describe the underlying methodology as well as the accompanying app for generating and manipulating phase diagrams.

### 2.2.1 Phase Diagram App

Phase diagrams have multiple applications in materials science. Traditional phase diagrams generated from experiments show not only stable phases but also delineate solubility limits and temperature dependence. In contrast, because MP currently only models materials at zero temperature and pressure and does not model solubility limits, the resulting phase diagrams might be more accurately referred to as phase *stability* diagrams (we use the terms interchangeably here). Nevertheless, such phase stability diagrams show the stable phases in a given chemical system as well as the relevant phase equilibria at various compositional ratios. One major application of such phase diagrams is to serve as a "reality check" for new hypothetical materials. If the energy of that proposed material is low enough to be on or nearly on the phase stability diagram, there is a higher probability that the material will be stable enough to be synthesized in the lab (Sun et al. 2016). Phase stability diagrams are also useful for identifying possible decomposition products that might compete with a target phase.

Generating such computational phase diagrams requires knowledge of the formation energies of all possible materials within a chemical system. For example, calculating a ternary phase diagram requires knowledge of the formation energies of all the relevant unary, binary, and ternary phases in that system. For a typical ternary system, calculating the energy for all known phases would require several dozen calculations. However, because the MP database already contains precomputed energies for most known inorganic compounds, one can now avoid running all these simulations and directly create reasonably complete phase diagrams using the MP data set.

Mathematically, the set of stable points on a phase diagram can be determined using the convex hull construction, which is a method of finding the minima as a function of $n$ degrees of freedom (Barber et al. 1996). By calculating the convex hull for the total energies of various calculated DFT energies, globally stable structures can be found as well as the various tie-lines that connect stable phases. The convex hull construction can be used to construct phase diagrams for an arbitrary number of components.

Many known compounds are not thermodynamically stable, i.e., they do not appear on phase stability diagrams (Sun et al. 2016). An additional metric is then necessary to distinguish the degree of metastability for these compounds. The construction of a convex hull provides an envelope of stability. Compounds

on the convex hull are stable, while compounds above the hull in energy are metastable. The energetic distance to the hull at the composition is thus a quantifiable metric and directly related to the metastability of that compound. A lower energy above the hull is typically desirable for synthesis because it implies less of an energy penalty to form the target compound compared to the known stable phase. Many of the known metastable compounds in the Materials Project are within 15 meV/atom of the hull, but depending on chemistry can extend past 60 meV/atom above the hull (Sun et al. 2016). While this analysis focuses on the metastability of known compounds, there is still work needed to quantify the limits of metastability.

Thus far we have described the formalism for closed systems, i.e., ones in which the stoichiometric ratio of elements is fixed, but the same formalism can be equally applied to open systems in which one or more elements are held at a fixed chemical potential rather than held to a fixed amount. For example, experiments may be carried out in air, which essentially serves as an infinite reservoir of atmospheric elements such as oxygen and nitrogen at particular chemical potentials. The same experiment under flowing argon gas would still represent an open system, but one in which the chemical potentials of those elements are greatly reduced. Thus, in environments that are open to a particular element, the relevant control variable is the chemical potential of that element ($\mu_i$) rather than its compositional value. The chemical potential is then treated as an external variable to obtain a grand potential phase diagram.

Users of MP need not be familiar with all the methodological details (Ong et al. 2008) of computational phase diagram construction to generate and use them. The MP web site allows users to simply type (or click on a visual periodic table) the elements for the system they are interested in. This will generate a phase diagram that will graphically display the phase diagram as well as a list of stable and metastable/unstable materials. Figure 6 shows a screenshot of the MP phase diagram app for a grand potential phase diagram for Li-Fe-P-O with an oxygen potential of $-5.288$ eV. Note that since the oxygen composition is prescribed by the potential, it doesn't exist as a degree of freedom in the phase diagram, collapsing the quaternary phase diagram into a ternary phase diagram (with a slider for controlling the oxygen chemical potential).

Other apps similarly make available powerful underlying methodologies to a broad audience. For example, similar to the grand potential phase diagram, Pourbaix diagrams are projections of global stability into potential-pH space to model electrochemical stability. A methodology for calculating such diagrams by utilizing experimentally measured free energies of aqueous ions and the calculated DFT energies for solid phases available in the Materials Project was previously developed (Persson et al. 2012). This methodology was incorporated into a "Pourbaix app" that allows users to simply select the chemical system of interest, elemental ratio, and concentration of ions in order to generate a familiar Pourbaix diagram that leverages the MP data set and that can be visually and interactively explored by the user. In addition, the stability of individual materials relative to the most stable decomposition product may be generated as a heatmap overlaid on the Pourbaix
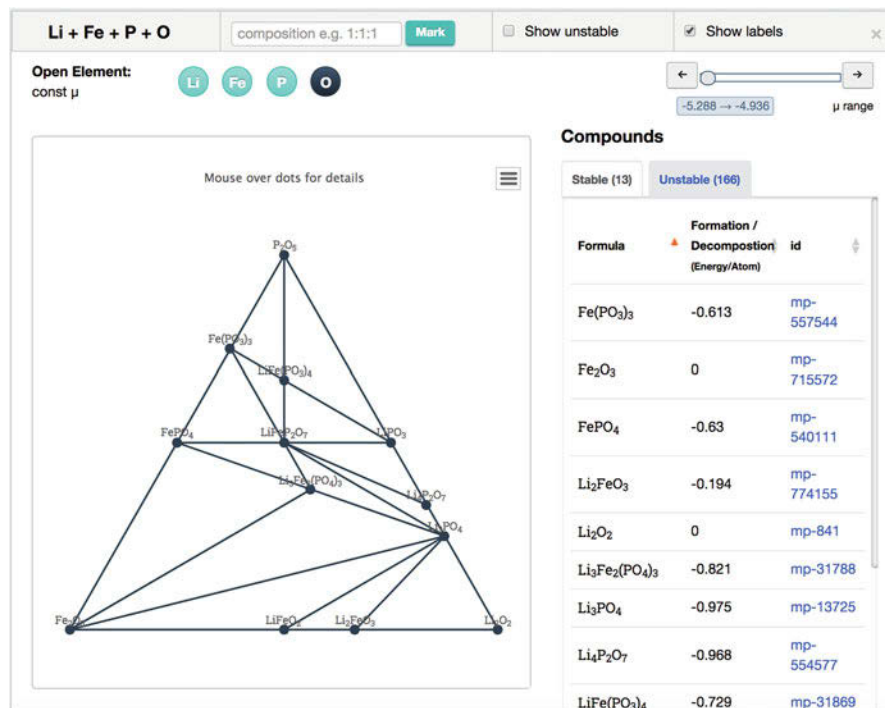
**Fig. 6** The Li-Fe-P-O ternary grand potential phase diagram open to oxygen as generated by the Materials Project's "phase diagram" app

diagram, providing users with a tool to estimate metastability under aqueous conditions (Singh et al. 2017).

## 3    Computation Infrastructure and Software Tools

Developing and maintaining calculation databases such as the Materials Project requires considerable attention to computing and software infrastructure. At the time of this writing, the Materials Project is the result of over one million individual calculations that represent over 100 million central processing unit (CPU) hours of computing time invested. Setting up, executing, analyzing, and managing all these calculations are far from straightforward. Here, we describe the infrastructure of the Materials Project at the time of this writing. However, we note that the economics of computing as well as the optimal choice of software libraries can change very quickly. The Materials Project infrastructure is therefore constantly evolving to apply the latest developments and best practices in computer science and software engineering to the field of materials science.

## 3.1    Computing Resources

The Materials Project (MP) has employed high-performance computing (HPC) resources at the US National Energy Research Scientific Computing Center (NERSC) and elsewhere, consuming over 100 million of CPU hours to date. Many-task computing workflows (Raicu et al. 2008) are increasingly using HPC environments because these resources typically offer the potential for large amounts of total computing time, good hardware specifications (e.g., moderate to high memory), and adequate storage. However, HPC environments present several challenges for running high-throughput calculations because these environments were originally designed to serve the needs of a small number of large, highly parallel applications that run for predictable times and perform all input/output to disk. In contrast, high-throughput calculations are typically extremely numerous and limited in achievable parallelism and require unpredictable, often very long total run times. In addition, they are often more suited to management by external services rather than solely through flat files on disk. To overcome these challenges, the Materials Project has developed a software library for running high-throughput calculations called "FireWorks" (Jain et al. 2015) that solves many of the computing challenges associated with running high-throughput jobs on HPC resources.

## 3.2    Choice of Database Software

Many portions of a high-throughput calculation workflow require efficient storage, retrieval, and search of information, including:

- Managing the state of high-throughput calculations
- Storage of the raw calculation results, and
- A searchable set of processed data for data dissemination and analysis

The Materials Project has chosen to use a not-only-SQL (NoSQL) "document store" (Cattell 2011), MongoDB, as its main database technology for these tasks (raw output files are also preserved). We note that this represents a shift from a other SQL-based data management strategies used previously in high-throughput computational materials science (Jain et al. 2011a). This decision was made primarily because MongoDB accommodates both the data heterogeneity and rapid pace of data model development required by the Materials Project. For example, unlike typical SQL relational database management systems (RDBMS) such as MySQL and PostgreSQL, MongoDB does not require a pre-designed, normalized schema between all data types at the beginning of the project. The types of data being stored continually evolve as we add new types of calculations into the project. By choosing MongoDB, MP can adapt quickly to these changes with small changes in application code instead of refactoring complex relational schemata.

Among document-oriented datastores, MongoDB is notable for its simple but powerful query language, ease of administration, and good performance on read-heavy workloads where most of the commonly accessed data (the so-called working set) can fit into memory. Its relative weaknesses for linking disparate data (database "joins") and write-heavy workloads are a reasonable trade-off for MP. A productivity benefit of MongoDB is that both the query language and the native data model are JavaScript Object Notation (JSON) (Bray 2017), which is the standard data format for modern web applications and easily represented and manipulated as native data types in the Python programming language (Van Rossum et al. 2007) in which our other software libraries are written. Thus, users familiar with Python and in particular its "dict" object can adapt quickly to understanding and developing data models with MongoDB. Our experience is that these aspects have allowed many more members of our team to collaborate on database development compared to our historical use of RDBMS in which only one or two members of the team were familiar enough with the system to make changes. More details on our experiences and challenges encountered in deploying a centralized datastore of this type within a scientific HPC ecosystem are described in Gunter et al. (2012).

## 3.3    Software Stack

### 3.3.1    Software to Perform and Analyze DFT Calculations

At the time of this writing, the Materials Project primarily uses density functional theory as implemented by the Vienna Ab Initio Simulation Package (VASP) (Kresse and Hafner 1994; Kresse and Furthmüller 1996). However, it is likely that other software packages such as ABINIT (Gonze et al. 2016) will play a larger role in MP in the future. Regardless of the choice of DFT implementation, the procedure for performing calculations involves many steps outside the core simulation. These steps include:

- Setting up the geometry for the material or system of interest
- Defining a workflow of calculations to compute the properties of interest
- Executing the calculations and correcting possible errors
- Analyzing, storing, and organizing the output data

The Materials Project has developed a comprehensive suite of software tools to accelerate and assist in these steps (see Fig. 7 for an overview).

Most of the compounds currently in the Materials Project use bulk crystal structure geometries as reported in the Inorganic Crystal Structure Database (ICSD) (Belsky et al. 2002). However, the computation of many properties requires performing algorithmic operations on these geometries. Examples include determining an appropriate ordered cell for sites with partial occupancies, creating appropriate slabs for surface calculations, and performing a series lattice deformations for computing elastic tensors. We have implemented routines for such geometry modifications in Python as part of the pymatgen (Ong et al. 2013) open-source software
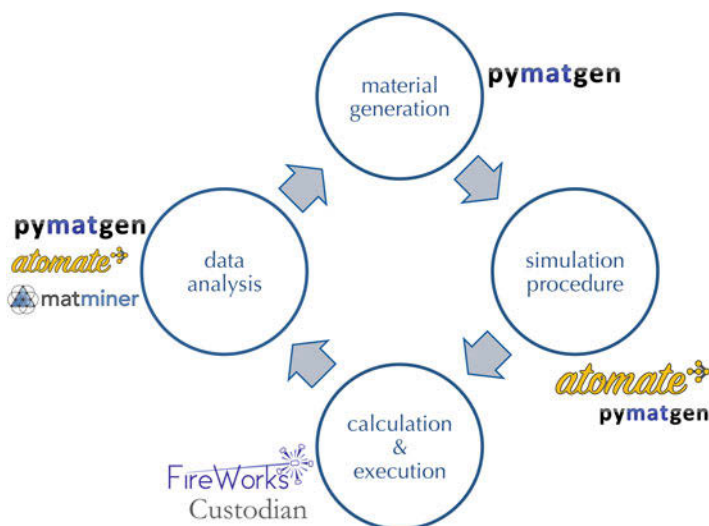
**Fig. 7** Various steps involved in data generation and analysis along with the relevant software stack for the Materials Project infrastructure

library. In many cases, these routines are directly implemented in pymatgen, whereas in others we provide an object-oriented Python wrapper to libraries released by the community such as spglib (Togo 2018) and enumlib (Hart and Forcade 2008). We note that the growth of web-based collaboration presents the opportunity for another method of generating new compounds: crowdsourced user suggestions. In this method, crystal structures designed by the user community (either offline or through Materials Project tools) are used as starting points for the calculation with the results reported back to the community. In its first three years of operation, this "MPComplete" service has been employed by over 800 unique users and has resulted in over 8,300 new materials added to MP's public database.

Depending on the property to be studied, a DFT "calculation" may in fact involve a series of individual computations that require data passing and modifications of geometry or input settings between computations. The set of calculations required for obtaining a desired output property, along with the dependencies and data passing requirements between these calculations, define a "workflow." The Materials Project has developed two software libraries in the Python programming language to manage such workflows. The first library, called "FireWorks" (Jain et al. 2015), is a general-purpose workflow library. FireWorks does not contain any materials science or DFT-specific code. Its scope is to provide a framework for users to define arbitrary sequences of calculations, store them in a database, execute them on various types of computing resource, and manage the status of potentially millions of workflows across systems. Thus, FireWorks is compatible with a broad class of scientific computing workflows (although it is best suited for high-throughput applications) and is frequently used outside the field of materials

science. The second library for workflow creation, "atomate" (Mathew et al. 2017), contains specific materials science workflows implemented in FireWorks and using pymatgen as a base library. The atomate package can be thought of as providing a library of materials science workflow implementations (e.g., standard workflows for electronic band structure, elastic properties, and piezoelectric, dielectric, and ferroelectric properties). Atomate users can specify an input geometry for a material and the desired workflow type, and atomate will provide a FireWorks-based implementation of that workflow that is ready to execute at supercomputing centers. Furthermore, atomate leverages the pymatgen library to automatically parse calculation outputs and create a database of materials properties that can be queried by the user. Various features in FireWorks and atomate allow for customization of behavior to specific situations, from low-level issues (such as interacting with various queueing systems) to high-level issues (such as running the same workflow with multiple DFT functional choices). Calculation workflows can also automatically adapt their procedure for later calculations based on the results obtained from earlier calculations.

When executing the calculation, it is possible to encounter various errors relating to calculation convergence. The Materials Project has developed a type of job wrapper to simulation software (e.g., VASP Kresse and Hafner 1994; Kresse and Furthmüller 1996 or QChem Kong et al. 2000) called "custodian" (Ong et al. 2014) that automatically monitors the output files of the calculation and automatically fixes errors (by stopping the job, changing the input files, and restarting the job) according to a set of rules. The custodian software can also be used to automate linear sequences of calculations (e.g., a convergence protocol that tightens numerical parameters until no change in output is achieved).

Once the calculation is executed, the results are parsed and stored in various database collections. Raw data is parsed by pymatgen as a component of atomate workflows. We note that pymatgen can parse output files (into structured data or as Python objects with callable functions) and can also perform high-level data analyses such as phase diagram creation or plotting. Separately, we employ code called "builders" that collect, reorganize, and post-process raw data into separate database collections that are more amenable to analysis than raw data collections. For example, a builder might collect together all calculated results on a single material to build a single summary report (a "material" document) for that compound. A builder might also collect together information from multiple compounds, perform an analysis, and store the results in a database. In service of such processes, we develop and use lightweight libraries to automate, simplify, and ultimately streamline the process of creating MongoDB databases. Our general "builder" code could be useful to any project that needs to perform extract-transform-load (ETL) operations with MongoDB. For example, they can be run in parallel without explicit coding of parallelism by the author. This allows CPU-intensive transformations of the data to run much faster on multi-core machines, which includes most modern hardware (integration with the Message Passing Interface (MPI) standard to enable parallelization across supercomputing resources is in development). Furthermore, facilities in our code for incremental building allow successive builds of source

MongoDB collection(s) to only operate on the records added since the last build, which can save significant amounts of computation time. Overall, the builder framework allows for efficient generation and reliable updating of multiple database collections that are tailored for different types of query and usage patterns.

### 3.3.2 Software for Data Dissemination: The Web Interface and RESTful API

The Materials Project places a strong emphasis on user experience, user interface design, and ensuring that data is efficiently disseminated so that a wide variety of users are able to apply the data for research, development, and education (Jain et al. 2016b). To this end, we have built an interactive web portal (https://www.materialsproject.org) focusing on the scientist as the end user. This web portal is built using the mature Django web framework (Dja 2015) due to its clean separation of front-end views from the back-end business logic. Django is written in the Python programming language, which eases integration with the pymatgen library and the growing scientific software ecosystem in Python. Django also provides a clear structure for organizing a so-called project into "apps," which maps well to our various interactive views across materials data such as compositional phase diagrams, Pourbaix diagrams, or domain-specific applications such as battery electrode searching. Additionally, Django features robust tools for user management, simplifying procedures for authentication (who someone is) and authorization (what a known someone can access/do). These tools are used, for example, to provide prepublication "sandboxes" for certain user groups within which to explore and perform analyses across private data sets prior to public release.

In order to tighten feedback loops for users searching data and using various functionalities that may not be applicable to all users, we organize our front-end code to asynchronously load both data and additional code using standard Asynchronous JavaScript And XML (AJAX) and Asynchronous Module Definition (AMD) protocols. Our choices of specific libraries for the web interface continue to evolve as trade-offs between established best practices (that are attractive for a system intended for continuous and reproducible use over many years) and emerging standards (that simplify ongoing maintenance and adding features).

Although many exploratory research studies are well suited to a graphical interface such as the one described above, other studies require programmatic access to this database. With this in mind, we have chosen to expose our data through an application programming interface (API) called the Materials API (MAPI) (Ong 2015). MAPI allows users to develop computer programs that can query, process, and download Materials Project data through a well-defined interface. To date, the MAPI has served more than 100 million requests for materials data for over 1500 distinct users.

APIs are used extensively throughout technology and software development. They serve to clearly and explicitly define a protocol for communicating with a piece of software or other system that is accessed programmatically. At the time of this writing, the most common framework for APIs that operate over the Internet
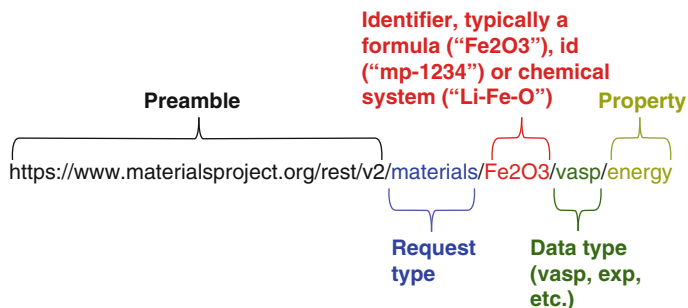
**Fig. 8** An example of the URL structure for the Materials API. (Reprinted from Ong (2015) with permission from Elsevier)

is REpresentational State Transfer (most commonly referred to as REST). The most simple use case for REST APIs is to map web uniform resource identifiers (URIs) to data (similar to how a computer's file system maps data to directories and filenames). In RESTful systems, information is organized into resources, each of which is uniquely identified via a uniform resource identifier (URI). In the case of Materials Project, each document or object (such as a computational task, crystal structure, or materials property) is represented by a URI (see Fig. 8 for an example) and an HTTP verb that can act on that object (GET, POST, PUT, DELETE, etc.). In most cases, this action returns structured data that represents the object, e.g., in the JavaScript Object Notation (JSON) format. For example, to request energy data (as calculated using VASP) on all $Fe_2O_3$ compounds in the Materials Project database, the URL shown in Fig. 8 could be constructed according to the protocol specified in the MAPI. We note that since MAPI is a RESTful system, users can interact with the MP database regardless of their computer system or programming language (as long as it supports basic HTTP requests.)

REST APIs allow for more powerful behavior to be seamlessly integrated alongside such basic information retrieval. For example, unique strings of characters associated with specific users (called API keys) can be used to manage access to resources. This is done by implementing the API in such a way that requires users to include their API keys in requests they make to the system and then implementing controls on the back end of the system to handle permissions and activity logging. RESTful APIs can also accept filtering parameters or other variables within requests to give users greater control over what they send or receive from a database. Moreover, URLs can be linked to more than just static resources; they can also point to back-end functions that enable interaction between a user program and MP. An example might be linking a URL such as "https://www.materialsproject.org/rest/v1/materials/snl/submit" to a function registering a request to compute a desired structure embedded in an http POST parameter.

Use of such an API offers a number of advantages. First, users do not have to be concerned with the actual architecture of the database they are interacting with or the details of its implementation since the API serves as a kind of "middleman" in

the process. From a user's perspective, the procedure to interact with the database is consistent over time, freeing the development team to make back-end changes without impacting the user's mode of interaction with the data. In addition, access to the database is system-agnostic. Anyone can develop an application in whatever environment they wish on top of the API with the confidence that it will be compatible with the MP database. Moreover, the data that users receive is always up-to-date, with no extra effort on their part, and the capabilities of the API can be seamlessly improved over time to give users access to even more powerful queries and analyses without creating new procedures for their use.

Although RESTful APIs can be intimidating to novices, they can be made more user-friendly by making the URL scheme explorable and hiding complexity through intermediate software layers. For example, a high-level Python interface to the MAPI called the MPRester is provided in the pymatgen (Ong et al. 2013) code base that allows users to obtain properties like crystal structure or electronic band structure using Python functions rather than explicit HTTP requests. We note that, whenever possible, the main Materials Project web site front end also avoids direct database queries and uses MAPI to query and access data in a way that is more maintainable and less prone to failure than custom interactions with the back-end software.

## 4    User Applications of the Materials Project to Research and Design Problems

Since its release, users of the Materials Project have used its data and tools in several hundred research studies (as highlighted in a previous review Jain et al. 2016a). In this section, we describe several recent examples and outline general strategies that have emerged in the literature for screening and designing materials for specific applications. While several of these studies involve active MP collaborators (Dagdelen et al. 2017; Yan et al. 2017; Chen et al. 2016; Zimmermann et al. 2017), a large fraction of the most recent studies that we found through a Web of Science search are from users that are not involved in Materials Project (Sendek et al. 2017; Shi et al. 2017; Ashton et al. 2017; Cheon et al. 2017; Choudhary et al. 2017; Lau et al. 2017; Shandiz and Gauvin 2016). This latter class of users perhaps most clearly demonstrates that it is possible to accelerate the research and design of new materials by generating and sharing materials information with the research community.

### 4.1    Phase Diagrams and Compound Stability

In studies that aim to improve our understanding based on experimental evidence or to synthesize new materials for a given application (Bayliss et al. 2014; Krishnamoorthy et al. 2015; Martinolich and Neilson 2014), generating phase equilibrium data is among the most frequently used MP capabilities . For example, Bayliss et al. conducted a study on a sodium-doped strontium silicate material

that had been claimed to feature remarkably high oxide ion conduction (Bayliss et al. 2014). By combining experiments (neutron powder diffraction, two-point AC impedance spectroscopy, time-of-flight secondary ion mass spectrometry) and DFT calculations, they could show that the conductivity was lower than previously reported and that the high energetic cost of oxygen vacancy formation was the underlying reason. Data from MP was used to cross-check the study's results for the DFT-PBEsol-derived major limiting phases of the $SrSiO_3$ decomposition.

Shi and co-workers (Shi et al. 2017) employed a high-throughput DFT screening approach for stable delafossite and related layered phases of composition $ABX_2$, where A and B are any elements from the periodic table and X a chalcogen (O, S, Se, and Te). From the initial materials set of 15,624 compounds, 285 were found to be within 50 meV/atom from the convex hull. While the majority of these structures are contained within the Materials Project database, the authors highlight that 79 of these stable systems are absent. This underscores that crystal structure databases such as MP still have considerable growth potential in terms of compound completeness and highlights the role that user-based compound submissions (through the MPComplete service, cf., Sect. 6.1) could play in extending such databases.

A similar example is the work by Krishnamoorthy et al. (2015), who used a high-throughput DFT-based screening to identify lead-free germanium iodide perovskites that could be used for light harvesting. The researchers computed the PBE bandgaps of 360 $AMX_3$ compositions, uncovering 9 interesting candidates. MP phase equilibrium data were used to further reduce the list by requiring that the materials be thermodynamically stable against decomposition to simpler binary phases. Three materials were left from the computational screening, $RbSnBr_3$, $CsSnBr_3$, and $CsGeI_3$, of which the latter was successfully synthesized and characterized. We refer the interested reader to a previous review article Jain et al. (2016b) for further examples of experimental studies conducted using MP-calculated phase diagrams.

## 4.2 Crystal Structure Analysis, 2D Materials, and Machine Learning

The large corpus of data available in the Materials Project can serve as a test bed for the development of new algorithms for processing of crystallographic data. This is the case for Ashton and co-workers (Ashton et al. 2017) who developed a topology-scaling algorithm to identify the dimensionality of a given crystal structure. They used the algorithm to search the MP database for materials that could be prospective 2D materials; 826 stable layered materials were identified, of which 680 were predicted to be feasible 2D material candidates based on the calculated exfoliation energy.

Similarly, Cheon and co-workers (Cheon et al. 2017) present an algorithm that can identify the dimensionality of weakly bonded subcomponents of a three-dimensional crystal structure. They apply this algorithm to >50,000 MP materials and identify 1,173 two-dimensional layered materials as well as 487 weakly bonded one-dimensional molecular chains, representing an order of magnitude increase

in the number of identified materials. Furthermore, 325 of these materials were suggested to be piezoelectric monolayers.

Interestingly, by specifically exploiting a weakness in typical DFT calculations that the dispersion forces are not well accounted for and thus lattice parameters of layered materials are often significantly inaccurate, Choudhary and co-workers (Choudhary et al. 2017) were able to identify two-dimensional material candidates. To this end, the authors required that the deviation between lattice constants from experiments and (mainly) MP database be ≥5%. In order to validate their approach, the authors used an accepted criterion based on the exfoliation energy and found that 88.9% of their predictions met this test.

Many structure-property relationships that form the basis of design rules in materials science are based on information pertaining to the local coordination environment. Therefore, it is highly desirable to have tools that effectively and efficiently identify basic local structural motifs such tetrahedra, octahedra, bcc, fcc, and hcp environments. Zimmermann et al. provided classification criteria for these motifs that are based on local structure order parameters, which were used to automatically identify these motifs in the entire Materials Project database. Additionally, these tools may also lead the way to alternative structure matching avenues (Zimmermann et al. 2017).

The abundance of data in the Materials Project also provides an opportunity to develop new machine learning (ML) techniques for modeling materials properties and for better understanding structure-property relationships. One such example of this appears in work from Faber et al. (2015) aimed at developing representations of periodic systems adaptable to ML models. In this study, 4000 structures from the Materials Project were used to evaluate the generalization error in the predicted formation energy based on three different crystal structure representation schemes and using kernel ridge regression, revealing that a sine matrix approach intended to simulate an infinite Coulomb sum was superior in its efficiency and accuracy.

Similarly, de Jong et al. (2016) demonstrated a machine learning approach to predicting elastic moduli of k-nary compounds that was effective over a highly diverse set of chemistries. More specifically, this study used gradient boosting machine local polynomial regression (GBM-Locfit) over the MP elastic tensor data set to determine a set of relevant descriptors and to derive elastic modulus predictions. Ultimately, this model was leveraged to estimate the Vickers hardness of the entire MP materials library, enabling a rapid search for superhard materials.

In most cases, easily retrievable or computable data such as the space group, composition, and the density are used in order to predict more complex properties such as the formation energy or the elastic tensor. Shandiz and Gauvin pursued the inverse route (Shandiz and Gauvin 2016): the authors conducted a classification study of 339 materials from the MP database that are potential Li-ion silicate cathodes (general composition: Li-Si-(Mn, Fe, Co)-O). In particular, they tested whether or not they could predict the crystal system (monoclinic, orthorhombic, or triclinic) based on features that were derived from both the input crystal structure and DFT outputs: the unit cell volume, the bandgap, the number of sites in the unit

cell, the formation energy, and the energy above the convex hull. Pair correlation plots of these features indicated that there was no exploitable direct correlation between any of the features and the crystal system. Decision tree-based methods (random forest and extremely randomized trees) were shown to yield prediction accuracies of up to 75%, and these methods performed better than linear and shrinkage discriminant analysis, respectively, artificial neural networks, support vector machines, and $k$-nearest neighbor classification.

## 4.3   Screening Materials for Applications

Perhaps the most consistent materials screening strategy that has emerged from the data on the Materials Project is that of filtering materials on successively tighter criteria appropriate to a given application space. In this approach, a filter common to most applications is typically on stability via $\Delta E_{hull}$, which can provide an indicator of whether a compound will be experimentally feasible. As illustrated in Fig. 9, successive filters in turn reduce the number of materials to be considered until it reaches a tractable quantity for follow-up with either more sophisticated calculations or for experimental inquiry.

This was the approach taken by Sendek and co-workers (Sendek et al. 2017) who searched for new candidate materials that could be used as solid-state electrolytes for lithium-ion batteries. The authors screened 12,831 Li-containing compounds from the Materials Project to filter those with high structural and chemical stability, low electronic conductivity, and low cost, thus, eliminating 92.2% of their initial
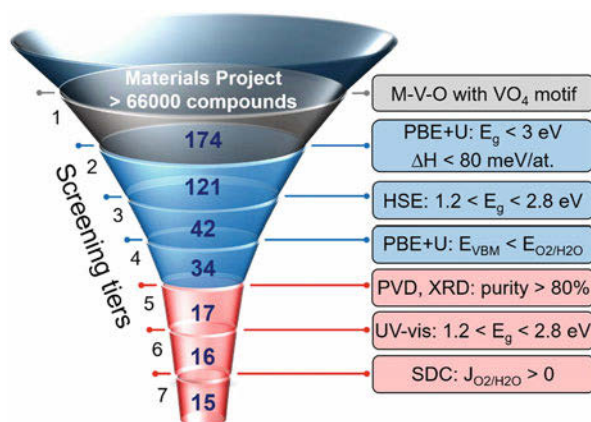


**Fig. 9** The "funnel" approach to materials screening through successive criterion filtering as applied to designing materials for solar fuel photoelectrocatalysis by Yan et al. (2017). Such approaches start with a large list of potential candidate materials and use a series of criteria (generally of increasing cost or complexity) to reduce the space of possibilities. (Reproduced from Yan et al. (2017); copyright 2017 National Academy of Sciences)

materials. Subsequently, an ionic conductivity classification model, which was trained on 40 crystal structures and associated measurements from literature, reduced the list of interesting candidates down to only 21 materials. In the latter step, the consideration of a multi-descriptor model over single-descriptor functions was critical to achieve predictive power. Many of the remaining 21 materials have not yet been studied experimentally, which hence offers new opportunities for experimental electrolyte research.

Identifying structurally similar compounds for the purpose of screening structure-sensitive properties and classifying materials has also begun to emerge as a screening strategy and design paradigm. Dagdelen and co-workers (Dagdelen et al. 2017) demonstrate such a screening procedure for predicting new auxetic materials (compounds with negative Poisson's ratios). The authors systematically screened the entire MP database via the Materials Project's REST API and compared each structure to $\alpha-$cristobalite $SiO_2$, one of the only inorganic crystalline materials previously known to exhibit a negative homogeneous Poisson's ratio. By coupling pymatgen's structure matching algorithm (which can match structures within a user-defined tolerance irregardless of crystal setting, supercell size, or composition) with more conventional screening strategies, 30 likely candidates were gleaned from over 65,000 structures. The full elastic tensor of each candidate was then calculated and their Poisson's ratios subsequently derived. Of these 30 structures, 3 were found to be homogeneously auxetic, and an additional 9 were found to exhibit near-zero homogeneous Poisson's ratio, with experimental confirmation ongoing.

An example of *in silico* screening with the Materials Project that has led to experimentally confirmed materials discovery was presented by Lau et al. (2017). These authors searched for promising chemical looping air separation (CLAS) materials in the MP database through successive criterion filtering ("funnel" approach). Specifically, the applied search filters included (i) restricting binary and ternary compounds, (ii) identifying compounds that can undergo oxidation reactions (at this step, the phase diagram app was employed), and (iii) restricting the temperature and oxygen partial pressure ranges in which the oxidation reactions would be carried out to sensible limits. The approach resulted in 5,501 tentative compounds and 20,861 relevant redox reactions. Since the reduction enthalpy and the gravimetric $O_2$ capacity (Fig. 10) did not reveal any exploitable trends, the authors had to employ a more heuristic route to reduce the candidate list. First, they required the reaction complexity and the total number of phases present in the reactions to be minimal, yielding 292 materials. Second, they decreased the number further to 108 by excluding compounds with expensive and toxic materials as well as reactions that involved non-oxides after reduction. From the remaining materials, they picked the $ABO_3$ perovskites because of their flexibility in oxygen stoichiometry without large structure changes and the ease of synthesizing perovskites in general. They subsequently synthesized and characterized $SrFeO_{3-\delta}$, which has emerged as a promising CLAS candidate due to its thermodynamic and excellent cycling stability as well as its resistance to carbonation over the temperatures of operation (Lau et al. 2017). Further examples of compound discovery with the Materials Project can be found in prior reviews (Hautier et al. 2012; Jain et al. 2016c).

**Fig. 10** An example of materials screening for chemical looping air separation application using MP data. Each data point represents a reduction reaction for each distinct compound with the largest $\mu_{O_2}$. The predicted reduction temperature at $p_{O_2} = 2.1 \times 10^4$ Pa for each reaction is plotted against $\Delta H_{reduction}$ (upper) and $O_2$ gravimetric capacity (lower). (Reproduced (Adapted or in part) from Lau et al. (2017) with permission of The Royal Society of Chemistry)
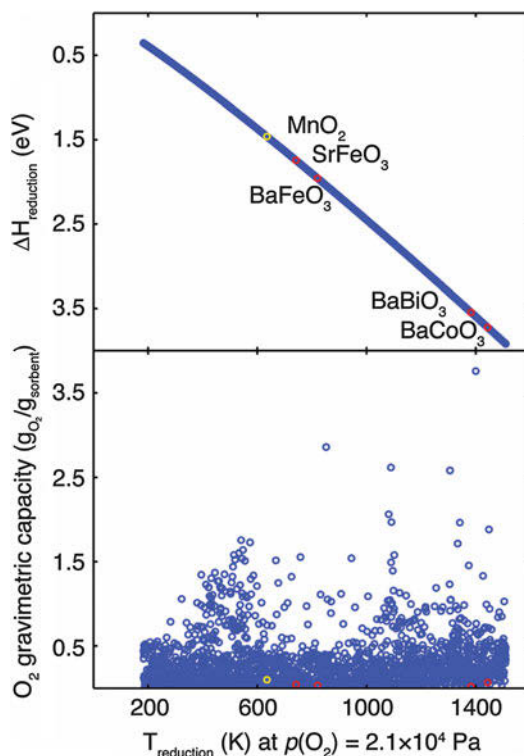


# 5 Outreach

Starting in 2016, the Materials Project has held annual workshops that have hosted more than 100 attendees from around the world. The workshops cover use of the Materials Project web site as well its software stack for performing and analyzing high-throughput calculations. Tutorials for the workshop utilized Jupyter (Ragan-Kelley et al. 2014) notebooks, which are a form of computer document that mixes formatted text, editable code, and interactive plots to illustrate a procedure. Participants were given the option to install the various codes to their own systems or to interact with a pre-installed environment configured using JupyterHub and Docker Swarm. The latter option allowed participants to focus on learning to use the software stack and left the details of individualized setup for later. All tutorials and course materials from these workshops are available online (Mathew et al. 2016; Winston et al. 2017).

Apart from the annual workshops, the Materials Project interacts with users in various ways. For example, MP has created YouTube videos with tutorials on all aspects of the web site, its various apps, and use of the API, which have had a total of over 30,000 views at the time of this writing. The Materials Project web site maintains a general-purpose discussion board

(https://discuss.materialsproject.org/) that has over 100 monthly active users, over 400 posts, and nearly 200 "likes" (whereby users quickly mark the helpfulness of posts) as of this writing. Finally, the MP software stack contains dedicated Google groups and Github issue pages where users and developers of the software can ask questions or get advice on software usage; hundreds of tickets have been resolved thus far.

# 6 Future of Materials Project

The advances in electronic structure theory, numerical algorithms, computing hardware, and software that have converged to make it possible to develop electronic structure databases are truly stunning. By leveraging these advancements, the Materials Project has computed millions of materials properties (e.g., electronic band structure, thermodynamic properties, mechanical properties, dielectric properties) across tens of thousands of materials, organized that information into searchable databases, and built rich web applications around the data in a way that would not have been possible a decade ago. The future efforts of the Materials Project will concentrate on further empowering the tens of thousands of scientists who design and develop new materials. Here, we describe some possible future developments to enhance property coverage, improve community data import capabilities, and provide an online materials design environment that leverages modern data analytics techniques.

## 6.1 Data Set Expansion

The Materials Project is continually generating new materials data at a rate of several tens of million CPU hours per year to expand the scope of its database. In the future, the Materials Project will expand in both breadth and depth: a greater variety of materials systems will be investigated, and more information will be calculated about individual materials. In terms of breadth, the Materials Project will expand to more completely encompass crystals with site disorder, i.e., partial site occupancies. The Materials Project will also continue its efforts and partnerships to expand its offerings of data on molecular, i.e., nonperiodic, systems. Finally, the Materials Project expects to play a more active role in not only computationally characterizing known materials but aiding experimentalists in the search for new materials yet to be discovered. In terms of depth, the Materials Project is expanding its library of computational workflows so that more information is available for each material in the database. Active areas of effort include phonon calculations and finite temperature properties, interfaces, spectroscopy, defects, and mapping relations between mechanical, thermal, and electrical effects. Furthermore, the Materials Project will leverage new advances in DFT functionals that make it possible to improve accuracy while still being computationally efficient for high-throughput computation.

This data set expansion will require orders of magnitude more computing resources than is currently employed. The Materials Project will continue to use "crowdsourcing," i.e., using input from the user community, as a method to prioritize various aspects of data set expansion. For example, the MPComplete service of Materials Project already allows users to both suggest new compounds for calculation and vote on compounds on which to prioritize more computationally expensive workflows. MPComplete then automatically integrates the results of each calculation with MP's core data set.

## 6.2    Beyond Simulations: Community-Contributed Materials Data

MP has become a worldwide resource for the materials sciences community, with over 40,000 users who rely on the portal as a trusted source to accelerate their research. This presents an opportunity to broaden the scope of MP's mission to also include assisting researchers disseminate their own data sets (whether computational or experimental) to the larger community of materials scientists. Thus, MP would serve not only as a hub for centrally generated computational data but would also host and distribute a variety of data sets generated by research groups worldwide. This will also give users of MP a more holistic picture of a compound because they would be presented with both computational and experimental information from a variety of techniques.

For this purpose, we soft-released our general contribution framework, MPContribs (Huck 2016b; Huck et al. 2015a, b, 2016), as a sustainable solution for well-curated data management, organization, and dissemination in the context of MP. Data as contributed through this framework as well as provenance and citation information for the contributors can be viewed on the MP web site. Early adopters are experimenting with MPContribs as a potential dissemination and hosting platform for their data, expanding the scope of data available through MP.

About a dozen early adopters have released landing pages to their contributed data sets on https://materialsproject.org/mpcontribs. Figure 11 highlights the landing pages for external studies of MnO2 phase selection, GLLB-SC bandgaps, dilute solute diffusion, and Fe-V-Co magnetic thin films. The last of these is based on data measured at the Advanced Light Source at Lawrence Berkeley National Laboratory, whereas the others are computationally derived. These landing pages can serve as interactive versions of the accompanying journal publications and allow research studies to be more easily reproduced and expanded upon.

## 6.3    MPCite: Citing Materials Data in Publications

The US Department of Energy Office of Scientific and Technical Information (OSTI) (Elliot et al. 2016) provides the E-Link service, which allows researchers to submit information about OSTI products (in form of XML meta-data records) and
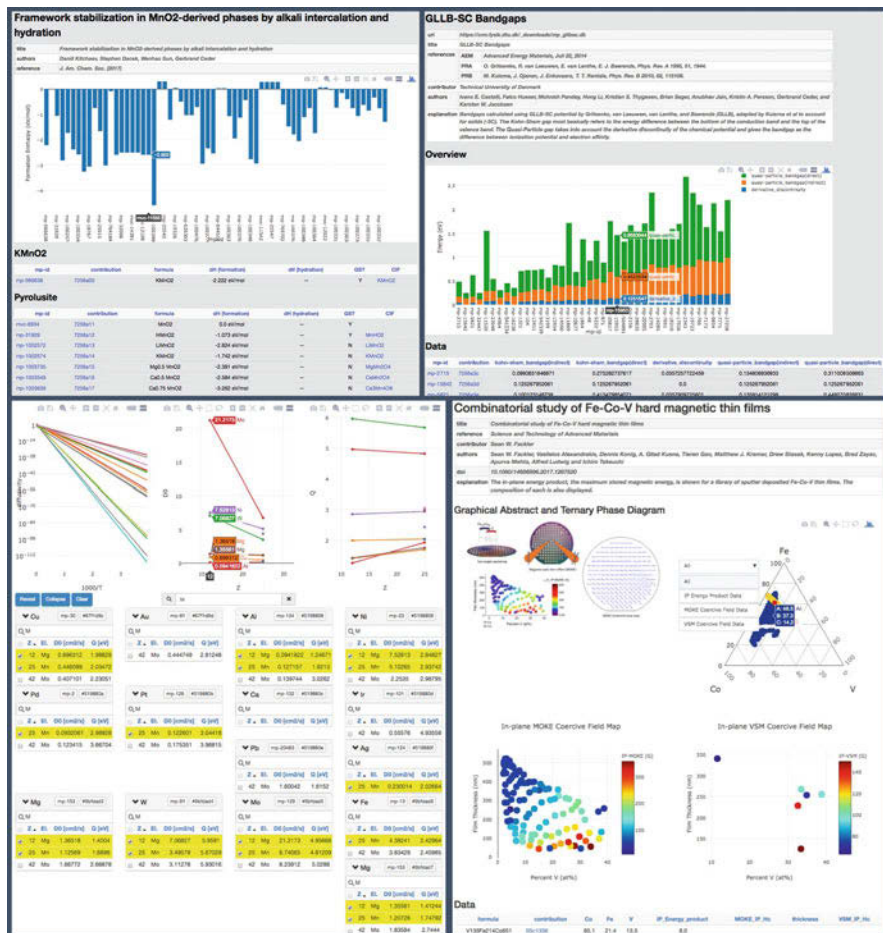
**Fig. 11** Examples of four different landing pages (representing different types of user-generated data sets) submitted to MPContribs

retrieve persistent digital object identifiers (DOIs) to identify it on the World Wide Web. DOIs are most commonly used for referencing and locating journal papers because they provide a unique URL linking to the journal's online landing page with more information about the publication. Our open-source software MPCite (Huck 2016a) enables the continuous request, validation, and dissemination of DOIs for all MP compounds. MPCite can also be employed for the assignment of DOIs to non-core database entries such as theoretical and experimental data contributed through MPContribs or user-generated analyses or structural data.

## 6.4    Data Analytics and Materials Design Environment

The Materials Project aims to not only generate raw data but also to empower users to make the best use of that data. For example, as described previously, we have found that many scientific studies conducted by users employ the "apps" built around the data such as phase diagram plots. As new data capabilities are established, we will continue to build additional apps to enable users to bridge the gap between a simple list of materials data and incorporating that data into a scientific analysis.

   The Materials Project will also place additional emphasis on helping users transform underlying data assets into new insights about structure-property relationships. In particular, new capabilities will allow users to formulate complex queries using visual interfaces and perform interactive data analysis and real-time filtering. Users will be able to rapidly iterate on materials design exploration with guidance provided by machine learning algorithms as well as traditional theory calculations. The four components of this vision for a materials design platform are Query, Process, Visualize, and Model/Compute (see Fig. 12). Next, we discuss these components in detail.

**Query**   Today, the Materials Project provides a visual web-based search interface to its underlying databases that is optimal for identifying a set of materials matching a series of constraints. However, many users require more sophisticated data pipelines in which one can visually add or remove filters and inspect the results at multiple points in the analysis or merge results from independent query streams. Such functionality is already possible for those that are capable of writing computer programs to fetch Materials Project data through MAPI, but remains difficult for others. New techniques of allowing users to fetch and interact with the data will be developed in the future so that one is able to call up exactly the desired data using a visual query interface.

**Process**   Once a user has compiled a data set of interest through the query tools, the Materials Project will make it easy for users to add descriptors/features to the data in a way that aids visualization, interpretation, and model building. We envision a system whereby a user can bring up any set of results (e.g., 100 materials of interest) and, by clicking a button, can rapidly generate a library of descriptors such as average electronegativity, local environment type, or polyhedral connection type for every material in the data set. Users will be able to use these descriptors to explore potential structure-property relationships through both conventional data analysis (e.g., visualization, statistical reports) and data mining and machine learning approaches.

**Visualize**   New software libraries and web frameworks such as Dash by Plotly and Crossfilter are making it easier than ever to produce high-quality charts on the web that can be interactively explored and manipulated. Such libraries can enable users
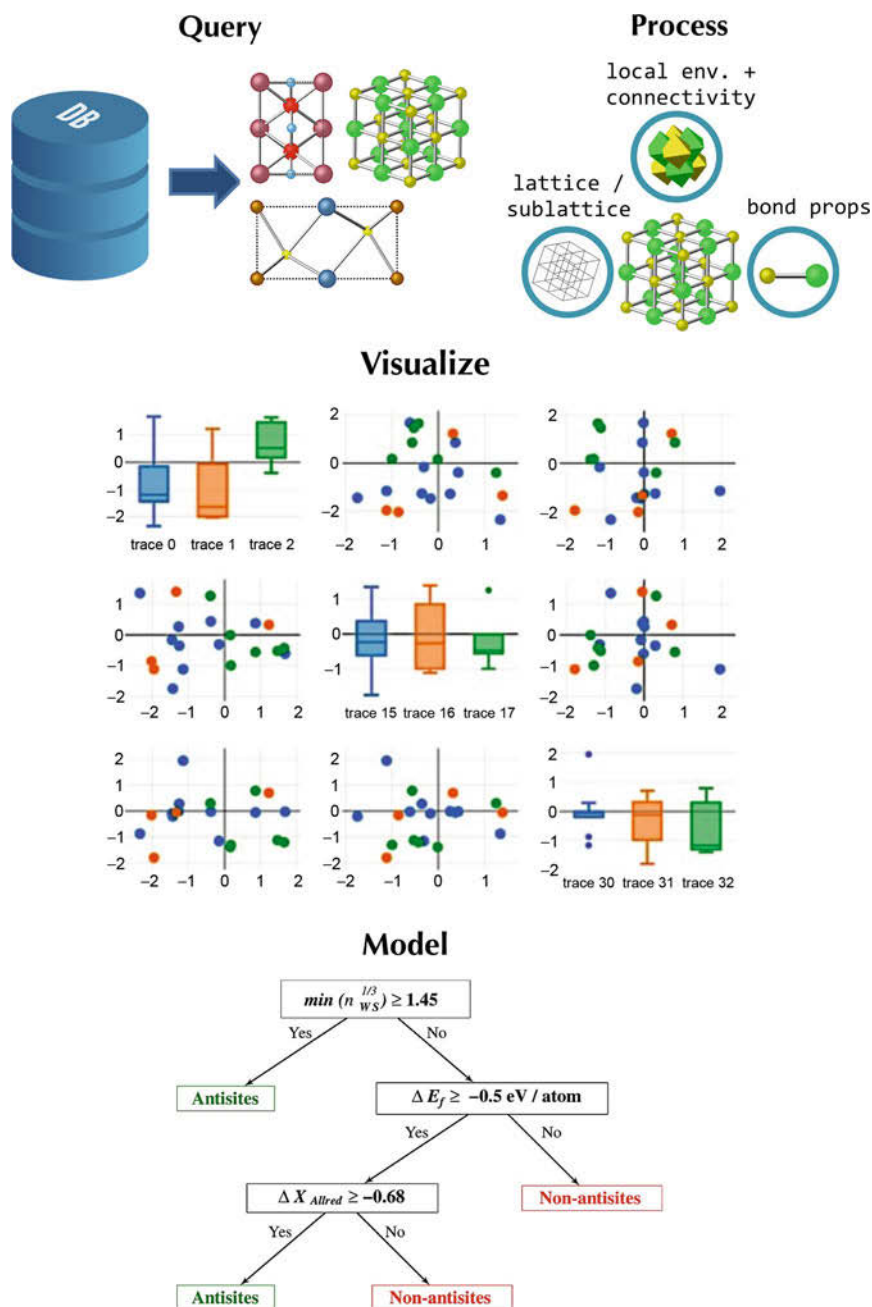
**Fig. 12** Four steps in data exploration and modeling for which MP is currently developing new features to assist the user. For a detailed description of these steps, see the main text

to perform a greater fraction of their data exploration on the MP web site itself. As a simple example, a user may decide to produce a standard X-Y scatterplot between two user-chosen properties of interest that additionally allows hovering over specific points to display details of that material. A more complex example would be to include tools for interactive filtering of the data set, with each modification or addition of a filter displaying live histogram charts of the distribution of various materials properties for the materials remaining in the data set. This will allow users to obtain immediate feedback on the distribution of various properties in their data set and visualize how various constraints and filters change this distribution.

**Model/Compute** After data preparation and exploration, the next step is to take action on the data. One possible action is to produce a model describing the various relationships between materials properties. For example, one may attempt to build a model that relates a structural descriptor such as local environment and a compositional descriptor such as redox active species to a computed output such as battery voltage. Machine learning models are an interesting way forward because, once trained and validated, they can be used to obtain nearly instantaneous feedback on how materials might behave even before any simulations are performed. Thus, such models can serve as surrogates for more complex and time-consuming physics simulations for qualitative estimation and ranking purposes. One can also imagine using these models to guide decisions regarding the computation of new materials.

With these elements in place, a single interactive web session would allow a user to perform sophisticated queries on the data set, automatically generate descriptors that could be useful in forming structure-property relationships, visually explore (and, if necessary, further refine) the data set, produce models that describe the data, and use those models to drive further computations. Indeed, many of these elements are present on the Materials Project today. For example, for materials in which elastic moduli are not yet computed, users can instantaneously obtain an estimate based on machine learning models (de Jong et al. 2016) as well as upvote the full computation based on density functional theory. In the future, this type of mixed usage of both data mining and conventional theory models will become more prevalent and increasingly natural to users.

## 6.5 Concluding Thoughts

Ab initio simulations have long been powerful tools for understanding and designing materials. With advances in high-throughput computing, it is now possible to create libraries of simulation results that can produce information on materials at a rate far surpassing that possible in the past. Furthermore, advances in software frameworks and web technologies have enabled the dissemination of these results in a barrier-free fashion to thousands of researchers worldwide. The Materials Project is an effort to make use of these advancements to build a valuable resource of materials data as well as software tools that transform the way materials are designed. In addition, the Materials Project aims to make computational materials science a more

collaborative process through the development of open-source software and through feedback from experimental groups.

It is an exciting time for theory – never before has there been so much materials data available or the potential of computation to make an impact in materials design higher. Experimentalists and theorists alike have been able to use the Materials Project to conduct scientific and industrial studies in a way that bridges traditional knowledge gaps. These use cases are likely an early sign of a future in which theoretical techniques and large materials databases will be increasingly influential and help to create a new materials design paradigm.

# References

Ashton M, Paul J, Sinnott SB, Hennig RG (2017) Topology-scaling identification of layered solids and stable exfoliated 2D materials. Phys Rev Lett 118:106101

Barber CB, Dobkin DP, Huhdanpaa H (1996) The quickhull algorithm for convex hulls. ACM Trans Math Softw 22(4):469–483. https://doi.org/10.1145/235815.235821

Bayliss RD, Cook SN, Scanlon DO, Fearn S, Cabana J, Greaves C, Kilner JA, Skinner SJ (2014) Understanding the defect chemistry of alkali metal strontium silicate solid solutions: insights from experiment and theory. J Mater Chem A 2:17919–17924

Belsky A, Hellenbrandt M, Karen VL, Luksch P (2002) New developments in the inorganic crystal structure database (ICSD): accessibility in support of materials research and design. Acta Crystall Sect B Struct Sci 58(3):364–369

Bray T (2017) The javascript object notation (JSON) data interchange format. STD 90, RFC 8259. https://www.rfc-editor.org/info/rfc8259

Cattell R (2011) Scalable SQL and NOSQL data stores. SIGMOD Rec 39(4):12–27. https://doi.org/10.1145/1978915.1978919

Chen W, Pohls JH, Hautier G, Broberg D, Bajaj S, Aydemir U, Gibbs ZM, Zhu H, Asta M, Snyder GJ, Meredig B, White MA, Persson K, Jain A (2016) Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment. J Mater Chem C 4:4414–4426

Cheon G, Duerloo KAN, Sendek AD, Porter C, Chen Y, Reed EJ (2017) Data mining for new two- and one-dimensional weakly bonded solids and lattice-commensurate heterostructures. Nano Lett 17:1915–1923

Choudhary K, Kalish I, Beams R, Tavazza F (2017) High-throughput identification and characterization of two-dimensional materials using density functional theory. Sci Rep 7:5179

Cococcioni M, de Gironcoli S (2005) Linear response approach to the calculation of the effective interaction parameters in the LDA + U method. Phys Rev B 71:035105. https://doi.org/10.1103/PhysRevB.71.035105

Dagdelen J, Montoya J, de Jong M, Persson K (2017) Computational prediction of new auxetic materials. Nat Commun 8:323

de Jong M, Chen W, Angsten T, Jain A, Notestine R, Gamst A, Sluiter M, Krishna Ande C, van der Zwaag S, Plata JJ, Toher C, Curtarolo S, Ceder G, Persson KA, Asta M (2015a) Charting the complete elastic properties of inorganic crystalline compounds. Sci Data 2:150009. https://doi.org/10.1038/sdata.2015.9, http://www.nature.com/articles/sdata20159

de Jong M, Chen W, Geerlings H, Asta M, Persson KA (2015b) A database to enable discovery and design of piezoelectric materials. Sci Data 2:150053. https://doi.org/10.1038/sdata.2015.53, http://www.nature.com/articles/sdata201553

de Jong M, Chen W, Notestine R, Persson K, Ceder G, Jain A, Asta M, Gamst A (2016) A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. Sci Rep 6:34256. https://doi.org/10.1038/srep34256, http://www.ncbi.nlm.nih.gov/pubmed/27694824, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5046120

Dja (2015) Django (version 1.8): the web framework for perfectionists with deadlines. https://djangoproject.com

Dozier A, Persson K, Ong SP, Mathew K, Zheng C, Chen C, Kas J, Vila F, Rehr J (2017) Creation of an XAS and EELS spectroscopy resource within the materials project using FEFF9. Microsc Microanal 23(S1):208–209

Elliot J, Vowell L, Nelson J, Ensor N, Robinson C, Studwell S, Martin M (2016) U.S. Department of Energy Office of Scientific and Technical Information (OSTI). https://www.osti.gov

Faber F, Lindmaa A, von Lilienfeld OA, Armiento R (2015) Crystal structure representations for machine learning models of formation energies. Int J Quant Chem 115(16):1094–1101. https://doi.org/10.1002/qua.24917

Gonze X, Jollet F, Araujo FA, Adams D, Amadon B, Applencourt T, Audouze C, Beuken JM, Bieder J, Bokhanchuk A, Bousquet E, Bruneval F, Caliste D, Côté M, Dahm F, Pieve FD, Delaveau M, Gennaro MD, Dorado B, Espejo C, Geneste G, Genovese L, Gerossier A, Giantomassi M, Gillet Y, Hamann D, He L, Jomard G, Janssen JL, Roux SL, Levitt A, Lherbier A, Liu F, Lukacevic I, Martin A, Martins C, Oliveira M, Poncé S, Pouillon Y, Rangel T, Rignanese GM, Romero A, Rousseau B, Rubel O, Shukri A, Stankovski M, Torrent M, Setten MV, Troeye BV, Verstraete M, Waroquier D, Wiktor J, Xue B, Zhou A, Zwanziger J (2016) Recent developments in the ABINIT software package. Comput Phys Commun 205:106. https://doi.org/10.1016/j.cpc.2016.04.003, http://www.sciencedirect.com/science/article/pii/S0010465516300923

Grindy S, Meredig B, Kirklin S, Saal JE, Wolverton C (2013) Approaching chemical accuracy with density functional calculations: diatomic energy corrections. Phys Rev B 87(7):075150

Gunter D, Cholia S, Jain A, Kocher M, Persson K, Ramakrishnan L, Ong SP, Ceder G (2012) Community accessible datastore of high-throughput calculations: experiences from the materials project. In: 2012 SC companion: high performance computing, networking storage and analysis, pp 1244–1251. https://doi.org/10.1109/SC.Companion.2012.150

Hart GL, Forcade RW (2008) Algorithm for generating derivative structures. Phys Rev B 77(22):224115

Hautier G, Jain A, Ong SP (2012) From the computer to the laboratory: materials discovery and design using first-principles calculations. J Mater Sci 47:7317–7340

Huck P (2016a) MPCite GitHub Repository. https://github.com/materialsproject/MPCite

Huck P (2016b) MPContribs GitHub Repository. https://github.com/materialsproject/MPContribs

Huck P, Gunter D, Cholia S, Winston D, N'Diaye A, Persson KA (2015a) User applications driven by the community contribution framework MPContribs in the materials project. http://arxiv.org/abs/1510.05727

Huck P, Jain A, Gunter D, Winston D, Persson KA (2015b) A community contribution framework for sharing materials data with materials project. http://arxiv.org/abs/1510.05024

Huck P, Jain A, Gunter D, Cholia S, Winston D, Persson K (2016) Materials project as analysis and validation hub for experimental and computational materials data. http://www.mrs.org/technical-programs/programs_abstracts/2016_mrs_fall_meeting_exhibit/tc2/tc2_5_3/tc2_5_06_6

Jain A, Hautier G, Moore CJ, Ong SP, Fischer CC, Mueller T, Persson KA, Ceder G (2011a) A high-throughput infrastructure for density functional theory calculations. Comput Mater Sci 50(8):2295–2310

Jain A, Hautier G, Ong SP, Moore CJ, Fischer CC, Persson KA, Ceder G (2011b) Formation enthalpies by mixing GGA and GGA + $U$ calculations. Phys Rev B 84:045115. https://doi.org/10.1103/PhysRevB.84.045115

Jain A, Ong SP, Chen W, Medasani B, Qu X, Kocher M, Brafman M, Petretto G, Rignanese GM, Hautier G, Gunter D, Persson KA (2015) Fireworks: a dynamic workflow system designed for high-throughput applications. Concurr Comput Pract Exp 27(17):5037–5059. https://doi.org/10.1002/cpe.3505

Jain A, Hautier G, Ong SP, Persson K (2016a) New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships. J Mater Res 31(08):977–994. https://doi.org/10.1557/jmr.2016.80, http://www.journals.cambridge.org/abstract_S0884291416000807

Jain A, Persson KA, Ceder G (2016b) Research update: the materials genome initiative: data sharing and the impact of collaborative ab initio databases. APL Mater 4(5):053102. https://doi.org/10.1063/1.4944683

Jain A, Shin Y, Persson KA (2016c) Computational predictions of energy materials using density functional theory. Nat Rev Mater 1:15004

Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133

Kong J, White CA, Krylov AI, Sherrill D, Adamson RD, Furlani TR, Lee MS, Lee AM, Gwaltney SR, Adams TR et al (2000) Q-chem 2.0: a high-performance ab initio electronic structure program package. J Comput Chem 21(16):1532–1548

Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Mater Sci 6(1):15–50. https://doi.org/10.1016/0927-0256(96)00008-0, http://www.sciencedirect.com/science/article/pii/0927025696000080

Kresse G, Hafner J (1994) Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. J Phys Condens Matter 6(40):8245–8257. http://iopscience.iop.org/article/10.1088/0953-8984/6/40/015

Krishnamoorthy T, Ding H, Yan C, Leong WL, Baikie T, Zhang Z, Sherburne M, Li S, Asta M, Mathews N, Mhaisalkar SG (2015) Lead-free germanium iodide perovskite materials for photovoltaic applications. J Mater Chem A 3:23829–23832

Lau CY, Dunstan MT, Hu W, Grey CP, Scott SA (2017) Large scale in silico screening of materials for carbon capture through chemical looping. Energy Environ Sci 10:818–831

Lejaeghere K, Bihlmayer G, Björkman T, Blaha P, Blügel S, Blum V, Caliste D, Castelli IE, Clark SJ, Dal Corso A, de Gironcoli S, Deutsch T, Dewhurst JK, Di Marco I, Draxl C, Dułak M, Eriksson O, Flores-Livas JA, Garrity KF, Genovese L, Giannozzi P, Giantomassi M, Goedecker S, Gonze X, Grånäs O, Gross EKU, Gulans A, Gygi F, Hamann DR, Hasnip PJ, Holzwarth NAW, Iuşan D, Jochym DB, Jollet F, Jones D, Kresse G, Koepernik K, Küçükbenli E, Kvashnin YO, Locht ILM, Lubeck S, Marsman M, Marzari N, Nitzsche U, Nordström L, Ozaki T, Paulatto L, Pickard CJ, Poelmans W, Probert MIJ, Refson K, Richter M, Rignanese GM, Saha S, Scheffler M, Schlipf M, Schwarz K, Sharma S, Tavazza F, Thunström P, Tkatchenko A, Torrent M, Vanderbilt D, van Setten MJ, Van Speybroeck V, Wills JM, Yates JR, Zhang GX, Cottenier S (2016) Reproducibility in density functional theory calculations of solids. Science 351(6280). https://doi.org/10.1126/science.aad3000, http://science.sciencemag.org/content/351/6280/aad3000

Martinolich AJ, Neilson JR (2014) Pyrite formation via kinetic intermediates through low-temperature solid-state metathesis. J Am Chem Soc 136:15654–15659

Mathew K, Ong SP, Winston D, Montoya J, Aykol M, Dwaraknath S, Huck P (2016) Assets for the 2016 materials project workshop. https://doi.org/10.5281/zenodo.1040432

Mathew K, Montoya JH, Faghaninia A, Dwarakanath S, Aykol M, Tang H, Heng Chu I, Smidt T, Bocklund B, Horton M, Dagdelen J, Wood B, Liu ZK, Neaton J, Ong SP, Persson K, Jain A (2017) Atomate: a high-level interface to generate, execute, and analyze computational materials science workflows. Comput Mater Sci 139(Supplement C):140–152. https://doi.org/10.1016/j.commatsci.2017.07.030, http://www.sciencedirect.com/science/article/pii/S0927025617303919

Ong SP (2015) The materials application programming interface (API): a simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. Comput Mater Sci 97:209–215. https://doi.org/10.1016/j.commatsci.2014.10.037, http://www.sciencedirect.com/science/article/pii/S0927025614007113

Ong SP, Wang L, Kang B, Ceder G (2008) Li–Fe–P–O$_2$ phase diagram from first principles calculations. Chem Mater 20(5):1798–1807

Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier VL, Persson KA, Ceder G (2013) Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. Comput Mater Sci 68:314–319. https://doi.org/10.1016/j.commatsci.2012.10.028, http://www.sciencedirect.com/science/article/pii/S0927025612006295

Ong SP, Qu X, Richards W, Dacek S, Jain A, Hautier G, Kitchaev D (2014) Custodian: a simple, robust and flexible just-in-time job management framework in python. https://doi.org/10.5281/zenodo.11714

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865

Perdew JP, Ernzerhof M, Zupan A, Burke K (1998) Nonlocality of the density functional for exchange and correlation: physical origins and chemical consequences. J Chem Phys 108(4):1522–1531

Persson KA, Waldwick B, Lazic P, Ceder G (2012) Prediction of solid-aqueous equilibria: scheme to combine first-principles calculations of solids with experimental aqueous states. Phys Rev B 85:235438. https://doi.org/10.1103/PhysRevB.85.235438

Petousis I, Mrdjenovich D, Ballouz E, Liu M, Winston D, Chen W, Graf T, Schladt TD, Persson KA, Prinz FB (2017) High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. Sci Data 4. https://www.nature.com/articles/sdata2016134

Ragan-Kelley M, Perez F, Granger B, Kluyver T, Ivanov P, Frederic J, Bussonnier M (2014) The Jupyter/Ipython architecture: a unified view of computational research, from interactive exploration to communication and publication. In: American Geophysical Union, Fall Meeting 2014, abstract id. H44D-07 https://ui.adsabs.harvard.edu/abs/2014AGUFM.H44D..07R/abstract

Raicu I, Foster IT, Zhao Y (2008) Many-task computing for grids and supercomputers. In: 2008 workshop on many-task computing on grids and supercomputers, pp 1–11. https://doi.org/10.1109/MTAGS.2008.4777912

Ricci F, Chen W, Aydemir U, Snyder GJ, Rignanese GM, Jain A, Hautier G (2017) Data descriptor: an ab initio electronic transport database for inorganic materials. Sci Data 4:170085

Sendek AD, Yang Q, Cubuk ED, Duerloo KAN, Cui Y, Reed EJ (2017) Holistic computational structure screening of more than 12,000 candidates for solid lithium-ion conductor materials. Energy Environ Sci 10:306–320

Shandiz MA, Gauvin R (2016) Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries. Comput Mater Sci 117:270–278

Shi J, Cerqueira TFT, Cui W, Nogueira F, Botti S, Marques MAL (2017) High-throughput search of ternary chalcogenides for p-type transparent electrodes. Sci Rep 7:43179

Singh AK, Zhou L, Shinde A, Suram SK, Montoya JH, Winston D, Gregoire JM, Persson KA (2017) Electrochemical stability of metastable materials. Chem Mater 29(23):10159–10167. http://pubs.acs.org/doi/abs/10.1021/acs.chemmater.7b03980

Sun W, Dacek ST, Ong SP, Hautier G, Jain A, Richards WD, Gamst AC, Persson KA, Ceder G (2016) The thermodynamic scale of inorganic crystalline metastability. Sci Adv 2:e1600225

Togo A, Tanaka I (2018) Spglib: a software library for crystal symmetry search. arXiv e-prints: 1808.01590. http://adsabs.harvard.edu/abs/2018arXiv180801590T

Tran R, Xu Z, Radhakrishnan B, Winston D, Sun W, Persson KA, Ong SP (2016) Surface energies of elemental crystals. Sci Data 3:160080. https://doi.org/10.1038/sdata.2016.80, http://www.nature.com/doifinder/10.1038/cgt.2016.38, http://www.nature.com/articles/sdata201680

Van Rossum G et al (2007) Python programming language. In: USENIX annual technical conference, vol 41, p 36

Wang L, Maxisch T, Ceder G (2006) Oxidation energies of transition metal oxides within the GGA + U framework. Phys Rev B 73:195107. https://doi.org/10.1103/PhysRevB.73.195107

Winston D, Mathew K, Montoya J, Huck P, Dwaraknath S, Dagdelen J, Liu M, Horton M, Jain A (2017) Assets for the 2017 materials project workshop. https://doi.org/10.5281/zenodo.1040436

Yan Q, Yu J, Suram SK, Zhou L, Shinde A, Newhouse PF, Chen W, Li G, Persson KA, Gregoire JM, Neaton JB (2017) Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment. Proc Natl Acad Sci 114(12):3040–3043. https://doi.org/10.1073/pnas.1619940114

Zhou F, Cococcioni M, Marianetti CA, Morgan D, Ceder G (2004) First-principles prediction of redox potentials in transition-metal compounds with LDA + $U$. Phys Rev B 70:235121. https://doi.org/10.1103/PhysRevB.70.235121

Zimmermann NER, Horton MK, Jain A, Haranczyk M (2017) Assessing local structure motifs using order parameters for motif recognition, interstitial identification, and diffusion path characterization. Front Mater 4:34

# The AFLOW Fleet for Materials Discovery

# 82

Cormac Toher, Corey Oses, David Hicks, Eric Gossett, Frisco Rose,
Pinku Nath, Demet Usanmaz, Denise C. Ford, Eric Perim,
Camilo E. Calderon, Jose J. Plata, Yoav Lederer, Michal Jahnátek,
Wahyu Setyawan, Shidong Wang, Junkai Xue, Kevin Rasch,
Roman V. Chepulskii, Richard H. Taylor, Geena Gomez, Harvey Shi,
Andrew R. Supka, Rabih Al Rahal Al Orabi, Priya Gopal,
Frank T. Cerasoli, Laalitha Liyanage, Haihang Wang, Ilaria Siloi,
Luis A. Agapito, Chandramouli Nyshadham, Gus L. W. Hart,
Jesús Carrete, Fleur Legrain, Natalio Mingo, Eva Zurek,
Olexandr Isayev, Alexander Tropsha, Stefano Sanvito,
Robert M. Hanson, Ichiro Takeuchi, Michael J. Mehl,
Aleksey N. Kolmogorov, Kesong Yang, Pino D'Amico,
Arrigo Calzolari, Marcio Costa, Riccardo De Gennaro,
Marco Buongiorno Nardelli, Marco Fornari, Ohad Levy,
and Stefano Curtarolo

## Contents

C. Toher (✉) · C. Oses · D. Hicks · E. Gossett · F. Rose · P. Nath · D. Usanmaz · D. C. Ford ·
E. Perim · C. E. Calderon · M. Jahnátek · W. Setyawan · S. Wang · J. Xue · K. Rasch ·
R. V. Chepulskii · G. Gomez
Department of Mechanical Engineering and Materials Science, Duke University, Durham,
NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA
e-mail: toherc@gmail.com

J. J. Plata
Department of Mechanical Engineering and Materials Science, Duke University, Durham,
NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

Departamento de Química Física, Universidad de Sevilla, Sevilla, Spain

**Abstract**

The traditional paradigm for materials discovery has been recently expanded
to incorporate substantial data-driven research. With the intent to accelerate

Y. Lederer · O. Levy
Department of Mechanical Engineering and Materials Science, Duke University, Durham,
NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Physics, NRCN, Beer-Sheva, Israel

R. H. Taylor
Department of Mechanical Engineering and Materials Science, Duke University, Durham,
NC, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Materials Science and Engineering, Massachusetts Institute of Technology,
Cambridge, MA, USA

H. Shi
Center for Materials Genomics, Duke University, Durham, NC, USA

A. R. Supka
Department of Physics and Science of Advanced Materials Program, Central Michigan
University, Mount Pleasant, MI, USA

R. Al Rahal Al Orabi
Department of Physics and Science of Advanced Materials Program, Central Michigan
University, Mount Pleasant, MI, USA

Solvay, Design and Development of Functional Materials Department, AXEL'ONE Collaborative
Platform – Innovative Materials, Saint Fons Cedex, France

P. Gopal
Department of Physics and Science of Advanced Materials Program, Central Michigan
University, Mount Pleasant, MI, USA

the development and the deployment of new technologies, the AFLOW Fleet for computational materials design automates high-throughput first-principles calculations and provides tools for data verification and dissemination for a broad community of users. AFLOW incorporates different computational modules to robustly determine thermodynamic stability, electronic band structures, vibrational dispersions, thermomechanical properties, and more. The AFLOW data repository is publicly accessible online at `aflow.org`, with more than 1.8

F. T. Cerasoli · L. Liyanage · H. Wang · I. Siloi · L. A. Agapito
Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

C. Nyshadham · G. L. W. Hart
Department of Physics and Astronomy, Brigham Young University, Provo, UT, USA

J. Carrete
Institute of Materials Chemistry, TU Wien, Vienna, Austria

F. Legrain
Universitè Grenoble Alpes, Grenoble, France
CEA, LITEN, Grenoble, France

N. Mingo
CEA, LITEN, Grenoble, France

E. Zurek
Department of Chemistry, State University of New York at Buffalo, Buffalo, NY, USA

O. Isayev · A. Tropsha
Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC, Chapel Hill, NC, USA
Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

S. Sanvito
School of Physics and CRANN Institute, Trinity College, Dublin, Ireland
Center for Materials Genomics, Duke University, Durham, NC, USA

R. M. Hanson
Department of Chemistry, St. Olaf College, Northfield, MN, USA

I. Takeuchi
Center for Nanophysics and Advanced Materials, University of Maryland, College Park, MD, USA
Department of Materials Science and Engineering, University of Maryland, College Park, MD, USA

M. J. Mehl
United States Naval Academy, Annapolis, MD, USA
Center for Materials Genomics, Duke University, Durham, NC, USA

A. N. Kolmogorov
Department of Physics, Binghamton University, State University of New York, Binghamton, NY, USA
Center for Materials Genomics, Duke University, Durham, NC, USA

million materials entries and a panoply of queryable computed properties. Tools to programmatically search and process the data, as well as to perform online machine learning predictions, are also available.

# 1    Introduction

The AFLOW Fleet is an integrated software infrastructure for automated materials design (Curtarolo et al. 2013) centered around the Automatic Flow (AFLOW) (Curtarolo et al. 2012a) framework for computational materials science. It features multiple scientific software packages, including the AFLOW high-throughput frame-

K. Yang
Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA

Center for Materials Genomics, Duke University, Durham, NC, USA

P. D'Amico
CNR-NANO Research Center S3, Modena, Italy

Dipartimento di Fisica, Informatica e Matematica, Universitá di Modena and Reggio Emilia, Modena, Italy

A. Calzolari
Center for Materials Genomics, Duke University, Durham, NC, USA

CNR-NANO Research Center S3, Modena, Italy

Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

M. Costa
Brazilian Nanotechnology National Laboratory (LNNano), CNPEM, Campinas, Brazil

R. De Gennaro
Dipartimento di Fisica, Universit'a di Roma Tor Vergata, Roma, Italy

M. Buongiorno Nardelli
Department of Physics and Department of Chemistry, University of North Texas, Denton, TX, USA

Center for Materials Genomics, Duke University, Durham, NC, USA
e-mail: Marco.BuongiornoNardelli@unt.edu

M. Fornari
Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Physics and Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, MI, USA
e-mail: forna1m@cmich.edu

S. Curtarolo (✉)
Center for Materials Genomics, Duke University, Durham, NC, USA

Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin-Dahlem, Germany
e-mail: stefano@duke.edu

work, the AFLOW$\pi$ (Supka et al. 2017) medium-throughput framework, and the PAOFLOW (Buongiorno Nardelli et al. 2017) utility for electronic structure analysis, along with the AFLOW.org data repository (Curtarolo et al. 2012b), its associated representational state transfer application programming interface (REST-API)) (Taylor et al. 2014), and the AFLUX Search-API (Rose et al. 2017). These elements are well integrated with one another: a Python+JSON (JavaScript Object Notation) interface connects AFLOW, AFLOW$\pi$, and PAOFLOW; and all software packages access the AFLOW.org repository via the REST-API and the Search-API.

Similar infrastructure has been developed by initiatives such as the Materials Project (Jain et al. 2013), NoMaD (Scheffler et al. 2014), OQMD (Saal et al. 2013), the Computational Materials Repository (Landis et al. 2012), and AiiDA (Pizzi et al. 2016). The Materials Project uses the `pymatgen` (Ong et al. 2013) Python-language data generation software infrastructure, and their repository is available at `materialsproject.org`. The Novel Materials Discovery (NoMaD) Laboratory maintains an aggregate repository available at `nomad-repository.eu`, incorporating data generated by other frameworks including AFLOW. The Open Quantum Materials Database (OQMD) (Saal et al. 2013) uses tools such as `qmpy` to generate their database, which can be accessed at `oqmd.org`. The Atomic Simulation Environment (ASE) (Bahn and Jacobsen 2002) is used to generate the Computational Materials Repository, available at `cmr.fysik.dtu.dk`. The ASE scripting interface is also used by the Automated Interactive Infrastructure and Database (AiiDA) framework available at `aiida.net`, which revolves around relational databases for its overall design and data storage. Additional materials design utilities include the High-Throughput-Toolkit (HTTK) and the associated Open Materials Database, `httk.openmaterialsdb.se`, as well as the Materials Mine database available at `www.materials-mine.com`, while computationally predicted crystal structures can be obtained from the Theoretical Crystallography Open Database at `www.crystallography.net/tcod/`.

The AFLOW Fleet employs density functional theory (DFT) to perform the quantum mechanical calculations required to obtain materials properties from first principles. These DFT calculations are carried out by external software packages, namely, the Vienna Ab initio Simulation Package (VASP) (Kresse and Hafner 1993; Kresse and Furthmüller 1996) in the case of AFLOW and QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017) in the case of AFLOW$\pi$. Results are stored in the AFLOW.org repository (Curtarolo et al. 2012b) and made freely available online via the `aflow.org` web portal, which is programmatically accessible and searchable via the AFLOW Data REST-API (Taylor et al. 2014) and AFLUX Search-API (Rose et al. 2017), respectively. The repository currently contains calculated properties for over 1.8 million materials entries, including both experimentally observed and theoretically predicted structures, and new results are continuously being added. This AFLOW data is successfully applied to (i) formulate descriptors for the formation of disordered materials such as metallic glasses (Perim et al. 2016), (ii) find new magnetic materials (Sanvito et al. 2017) and superalloys (Nyshadham et al. 2017), (iii) generate phase diagrams for alloy systems (Barzilai et al. 2016, 2017a; Lederer et al. 2018) and identify new ordered compounds (Levy et al. 2010a, b, c;

Barzilai et al. 2017b), and (iv) train machine learning models to identify potential superconductors (Isayev et al. 2015) and predict electronic and thermomechanical properties (Isayev et al. 2017).

## 2    AFLOW: Efficient Data Generation

The AFLOW framework for computational materials science automates the full workflow for materials properties calculations (Curtarolo et al. 2012a). Using a standard set of calculation parameters (Calderon et al. 2015), input files are automatically generated for the VASP (Kresse and Hafner 1993; Kresse and Furthmüller 1996) DFT software package with projector-augmented-wave pseudopotentials (Blöchl 1994) and the PBE parameterization of the generalized gradient approximation to the exchange-correlation functional (Perdew et al. 1996). Calculations are monitored as they run to detect and correct for errors without the need for any user intervention. Useful materials data is then extracted and processed for dissemination through the AFLOW.org repository. The entire framework is written in the C++ programming language (more than 400,000 highly integrated lines, as of version 3.1.153), providing a robust platform for continuous infrastructure development with reliable high performance.

### 2.1    AFLOW: Automated Workflows

AFLOW offers several automated workflows, each dedicated to a specific type of characterization yielding a set materials properties. For electronic properties, AFLOW performs four DFT calculations: two rounds of geometry relaxation (stage name: "RELAX") using the VASP conjugate gradient optimization algorithm, a static run (i.e., fixed geometry; stage name: "STATIC") with a denser $\mathbf{k}$-point mesh to obtain an accurate density of states, and a band structure calculation (stage name: "BANDS") following the AFLOW Standard path through the high-symmetry $\mathbf{k}$-points in the Brillouin zone (Setyawan and Curtarolo 2010).

Other workflows in AFLOW manage ensembles of DFT calculations, all offering the same automated error-correction procedures for high-throughput processing. For thermal and elastic properties, the Debye-Grüneisen model (Automatic GIBBS Library, AGL) (Toher et al. 2014) is combined with the Automatic Elasticity Library (AEL) (Toher et al. 2017) as described in Sect. 2.6. A more accurate thermal characterization can be resolved with the finite displacement method for phonon calculations (Automatic Phonon Library, APL) (Nath et al. 2016) and its associated extensions, i.e., the quasi-harmonic approximation (QHA-APL) (Nath et al. 2016) and Automatic Anharmonic Phonon Library (AAPL) (Plata et al. 2017), as described in Sect. 2.7. AFLOW also extends beyond ideal crystalline materials characterization, offering modules to investigate off-stoichiometric materials (AFLOW-POCC, Sect. 2.5) (Yang et al. 2016) and to predict metallic glass formation as a function of composition (Perim et al. 2016).

## 2.2    AFLOW: Prototype Library

The AFLOW framework uses decorated crystal structure prototypes for materials discovery (Mehl et al. 2017). Structural prototypes are specific arrangements of atoms which are commonly observed in nature, such as the rock salt, zinc blende, and wurtzite structures. The atomic sites in these prototypes are populated with different elemental species to generate materials structures, for which the properties and thermodynamic stability are then obtained from DFT calculations. An extensive list of the structural prototypes included in AFLOW has been published in Mehl et al. (2017) and is available online at `http://aflow.org/CrystalDatabase`.

   Pages within the website display a curated list of data for each structural prototype, including materials exhibiting this structure, various symmetry descriptions, the primitive and atomic basis vectors, and original references where the structure was observed. Accompanying these descriptions is an interactive `Jmol` visualization of the prototype, as described in Sect. 2.8. The page also contains a prototype generator, where the structural degree(s) of freedom and atomic species are defined to create new materials by leveraging the AFLOW prototypes module. This generates the corresponding input file for one of many ab initio software packages, including VASP (Kresse and Hafner 1993; Kresse and Furthmüller 1996), QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017), ABINIT (Gonze et al. 2002), and FHI-AIMS (Blum et al. 2009).

## 2.3    AFLOW-SYM: Symmetry Analyzer

The AFLOW framework automatically analyzes the symmetry of materials structures and returns a complete symmetry description. To address numerical tolerance issues, AFLOW employs an atom mapping procedure that is reliable even for non-orthogonal unit cells and uses an adaptive tolerance scheme to ensure symmetry results are commensurate with crystallographic principles (see Fig. 1). These routines—referred to as AFLOW-SYM (Hicks et al. 2018)—are robust and have been used to successfully determine the symmetry properties of over 1.8 million materials in the AFLOW repository.

   Structural isometries are identified by determining the set of symmetry operators that lead to isomorphic mappings between the original and transformed atoms. The structure exhibits symmetry under a particular operation if the set of closest mapping distances are all below a tolerance threshold $\epsilon_0$. Periodic boundary conditions introduce complexity for finding the minimum mapping vector, necessitating the exploration of neighboring cells. This is achieved via the method of images through either (i) a unit cell expansion, yielding the globally optimal distance or (ii) a bring-in-cell method (generally performed in fractional coordinates) that reduces each component of the distance vector independently. While computationally inexpensive compared to the unit cell expansion, the bring-in-cell method is only exact for orthogonal lattices (i.e., described by a diagonal metric tensor), since it does not consider overlap between lattice vectors (see Fig. 1a). To safely exploit
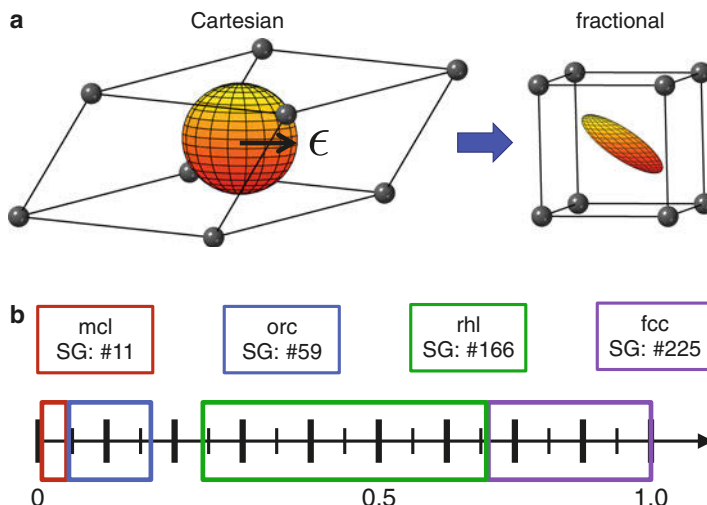
**Fig. 1 Visualization of tolerance-sphere warping and adaptive tolerance method.** (**a**) Illustration of the warping of space when transforming from cartesian to fractional coordinates in the general case. (**b**) Spectrum of space groups identified by AFLOW-SYM with different tolerance choices

the bring-in-cell approach, AFLOW-SYM employs a heuristic maximum tolerance $\epsilon_{\max}$ based on the maximum lattice skewness with a threshold which guarantees consistent and accurate results (Hicks et al. 2018).

Given a particular tolerance value, different symmetry operations can be realized in or excluded from the description of a crystal. Figure 1b highlights how the tolerance value affects the possible space groups for AgBr (ICSD #56551 with a reported space group #11). The neighboring space group regions are consistent with non-isomorphic subgroup relations, namely, between space groups #59 and #11 and between #225 and #166. However, a gap or "confusion" tolerance region occurs between space groups #59 and #166 (with no direct subgroup relations). The problematic regions stem from noise in the structural data, impeding the identification of operations consistent with symmetry principles. This problem is solved by using a radial tolerance scan extending from the input tolerance $\epsilon_0$. Given a change in tolerance, the algorithm recalculates and verifies all symmetry properties until a globally consistent description is identified.

AFLOW-SYM is compatible with many established ab initio input files, including those for VASP (Kresse and Furthmüller 1996), QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017), ABINIT (Gonze et al. 2002), and FHI-AIMS (Blum et al. 2009). From the structural information, AFLOW-SYM delivers the symmetries of the lattice, crystal (lattice + atoms), reciprocal lattice, superlattice (equally decorated sites), and crystal-spin (lattice + atoms + magnetic moment). This affords a multitude of symmetry descriptions to be presented, such as the space group number/symbol(s), Pearson symbol, point group symbol(s), Wyckoff positions, and Bravais lattice

type/variation (Setyawan and Curtarolo 2010). Moreover, the operators of the different symmetry groups—including the point groups, factor groups, space group, and site symmetries—are provided to users in rotation matrix, axis-angle, matrix generator, and quaternion representations for easy manipulation. All symmetry functions support the option to output in JSON format. This allows AFLOW-SYM to be leveraged from other programming languages such as Java, Go, Ruby, Julia, and Python—facilitating the incorporation of AFLOW-SYM into numerous applications and workflows.

## 2.4   AFLOW-CHULL: Convex Hull Analysis

Structure and energy data from the AFLOW.org repository are used to resolve the low-temperature/low-pressure thermodynamic stability of compound systems. For a given stoichiometry, the AFLOW.org repository provides the DFT energies of various crystal polymorphs. By exploring representative structures over the full range of stoichiometries, AFLOW-CHULL (Oses et al. 2018) constructs the minimum energy surface, i.e., the lower-half convex hull (Barber et al. 1996), defining thermodynamic stability for the system (at zero temperature and pressure). Structures on the hull are thermodynamically stable (ground state), while those far from the hull will decompose into a combination of stable phases, dictating synthesizability at these conditions. Any analysis of the hull requires sufficient statistics to ensure convergence, i.e., enough representative structures have been included in the alloy system calculations such that any additional entries are not expected to change the minimum energy surface.

The geometric construction offers several key properties critical for synthesizability. For a specific composition, the energetic distance to the hull quantifies the energy released during the decomposition, while the ground state phases defining the tie-line/facet below the compound are the products of the reaction. The distance from the hull also measures the "severity" of instability, i.e., structures near the hull may stabilize at higher temperatures or pressures. Similarly, a robust stability criterion can be quantified for ground state phases by removing the phase from the set and measuring the distance of the compound from the new hull. The larger the distance, the less likely the ground state phase will become unstable at higher temperatures/pressures (Sanvito et al. 2017). The generalized tie-lines (facet ridges) dictate which phases can coexist in equilibrium and play a role in determining the feasibility of synthesis/treatment techniques, such as precipitation hardening (Nyshadham et al. 2017).

Given a compound system, AFLOW-CHULL automatically queries the AFLOW.org database, constructs the hull, calculates the aforementioned properties, and delivers the information in one of the following formats: PDF, plaintext, and JSON. AFLOW-CHULL can also visualize the 2D and 3D hulls, as illustrated in Fig. 2. In the case of the PDF output format, hyperlinks are included to allow for additional queries of the full properties set offered through the AFLOW.org repository. Links are also
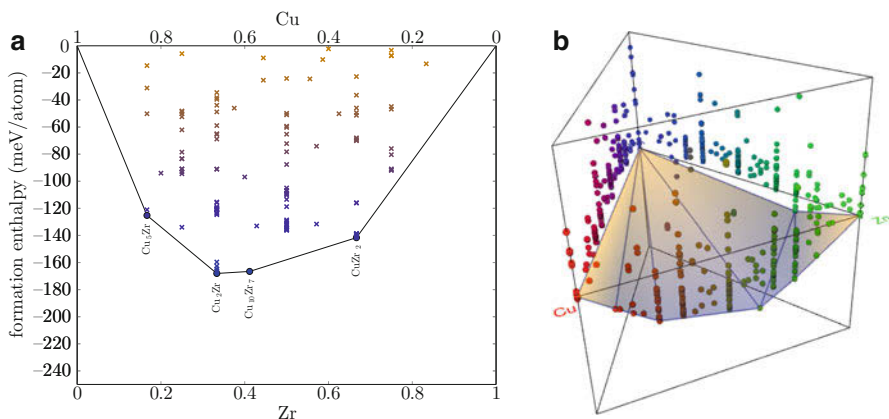
**Fig. 2 Example convex hull illustrations offered by AFLOW.** (**a**) 2D convex hull of the CuZr system generated automatically by AFLOW. (**b**) 3D convex hull of the CuMnZn system presented through the AFLOW-CHULL application online: http://aflow.org/aflow-chull

added connecting the hull visualization to relevant properties for easy navigation of the full PDF document.

A separate online application, available at aflow.org/aflow-chull, has been created to showcase the results of AFLOW-CHULL and provides interactive binary and ternary convex hull visualizations. The application consists of four components: the periodic table, visualization viewport, selected entries list, and the comparison page. The periodic table component is the entry point of the application and provides the interface to search for convex hulls of different alloy systems. Elements within the periodic table respond when selected to display information to the user. As a selection is made, the color of each border will change to green, yellow, and red based on hull reliability. A reliability threshold of 200 compounds for a binary hull has been heuristically defined. Selections highlighted in green are well above this threshold, while those in yellow/red are near/below the cutoff.

When a hull is selected, the application transitions to the visualization viewport component. Depending on the number of elements selected, a 2D plot (binary) or 3D plot (ternary) will appear. Each plot is interactive, allowing points to be selected, where each point represents an entry in the AFLOW repository. Information for each point is displayed in the selected entries list component, which is accessible through the navigation bar. Selected hulls will appear on the comparison page as a grid of cards, and selected points are highlighted across all hulls containing those entries.

## 2.5    AFLOW-POCC: Partial Occupations

The AFLOW Partial Occupation module (AFLOW-POCC) (Yang et al. 2016) models configurational and structural disorder including substitutions, vacancies, and
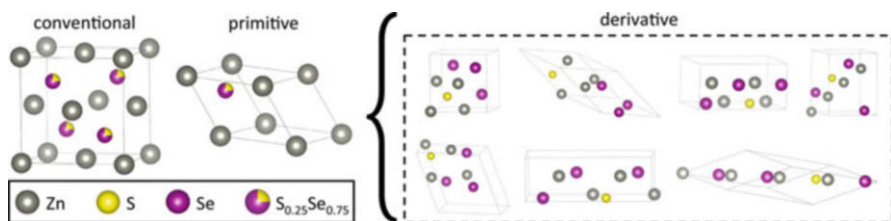
**Fig. 3 Structure enumeration for off-stoichiometric materials modeling.** For the off-stoichiometric material $ZnS_{0.25}Se_{0.75}$, a superlattice of size $n = 4$ accommodates the stoichiometry exactly. By considering all possibilities of decorated supercells and eliminating duplicates by UFF energies, seven structures are identified as unique. These representative structures are fully characterized by AFLOW and VASP and are ensemble-averaged to resolve the system-wide properties

random lattice site occupation, by generating a set of representative configurations. First, a supercell size is determined that accommodates the fractional stoichiometry to within a user-defined tolerance. Given the supercell size, $n$, superlattices are generated using Hermite Normal Form matrices (Hart and Forcade 2008), which are then decorated in accordance with the stoichiometry to generate all possible configurations, as illustrated in Fig. 3. Duplicate configurations are rapidly identified and eliminated by estimating the energy of each structure using the Universal Force Field (UFF) model (Rappe et al. 1992). The properties of the remaining unique configurations are calculated with DFT, and ensemble-averaged to resolve system-wide properties of the disordered material. The ensemble-average employs a Boltzmann distribution weight which is a function of a disorder parameter (temperature), energy relative to the ground state configuration, and degeneracy as determined by the UFF model. Ensemble-average properties include the electronic band gap, density of states, and magnetic moment.

## 2.6    AEL and AGL: Thermomechanical Properties

The AFLOW Automatic Elasticity Library (AFLOW-AEL Toher et al. 2017) and the AFLOW Automatic GIBBS Library (AFLOW-AGL Toher et al. 2014) modules determine thermomechanical materials properties from calculations of strained primitive cells. These methods are generally computationally less costly than the phonon (APL and AAPL) calculations described in Sect. 2.7, although APL and AAPL generally give more quantitatively accurate results, particularly for properties where anharmonic effects are important. AEL and AGL have been combined into a single automated workflow, which has been used to calculate the thermomechanical properties for over 5000 materials in the AFLOW repository.

The AEL module applies a set of independent normal and shear strains to the primitive cell of a material (de Jong et al. 2015; Toher et al. 2017) as depicted in Fig. 4a and uses DFT to calculate the resulting stress tensors. This set of strain-stress data is used to generate the elastic stiffness tensor, i.e., the elastic constants:
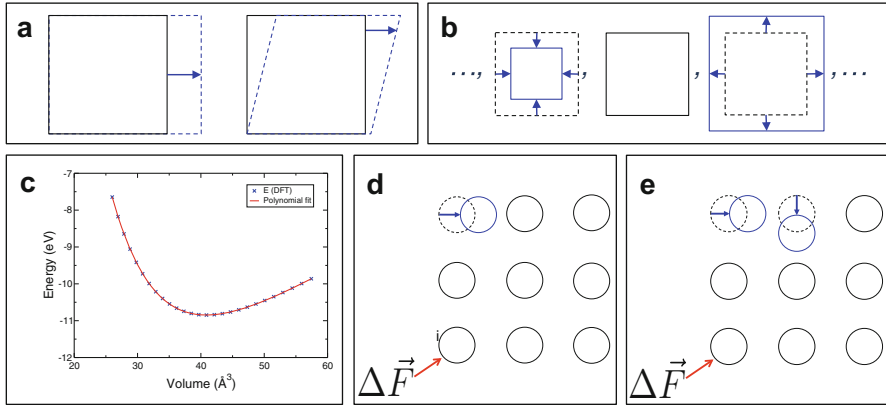
**Fig. 4** **Ensembles of distorted geometries for the calculation of thermomechanical properties with AFLOW.** AEL uses the stresses from a set of (**a**) normal and shear strained cells to obtain the elastic constants. AGL calculates the energies of a set of (**b**) isotropically compressed and expanded unit cells and (**c**) fits the resulting $E(V)$ data by a numerical polynomial or by an empirical equation of state. APL obtains the (**d**) second-order harmonic IFCs from a set of single-atom displacements and the (**e**) third-order anharmonic IFCs from a set of two-atom displacements

$$
\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{12} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{13} & c_{23} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{14} & c_{24} & c_{34} & c_{44} & c_{45} & c_{46} \\ c_{15} & c_{25} & c_{35} & c_{45} & c_{55} & c_{56} \\ c_{16} & c_{26} & c_{36} & c_{46} & c_{56} & c_{66} \end{pmatrix} \begin{pmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ 2\epsilon_{23} \\ 2\epsilon_{13} \\ 2\epsilon_{12} \end{pmatrix} \tag{1}
$$

written in the $6 \times 6$ Voigt notation using the mapping (Poirier 2000): $11 \mapsto 1$, $22 \mapsto 2$, $33 \mapsto 3$, $23 \mapsto 4$, $13 \mapsto 5$, $12 \mapsto 6$. These elastic constants are combined to calculate the bulk, $B$, and shear, $G$, elastic moduli in the Voigt, Reuss, and Voigt-Reuss-Hill (VRH, $B_{\mathrm{VRH}}$ and $G_{\mathrm{VRH}}$) approximations. The Poisson ratio $\nu$ is then given by:

$$
\nu = \frac{3B_{\mathrm{VRH}} - 2G_{\mathrm{VRH}}}{6B_{\mathrm{VRH}} + 2G_{\mathrm{VRH}}}. \tag{2}
$$

The AGL module is based on the GIBBS (Blanco et al. 1996, 2004) quasi-harmonic Debye-Grüneisen model and calculates the energy as a function of volume, $E(V)$, for a set of isotropically compressed and expanded strains of the primitive cell, as illustrated in Fig. 4b. The $E(V)$ data are fitted by either a numerical polynomial or an empirical equation of state to obtain the adiabatic bulk modulus $B_{\mathrm{S}}(V)$, as shown in Fig. 4c. The Debye temperature $\theta_{\mathrm{D}}(V)$ as a function of volume is then calculated using the expression:

$$\theta_D = \frac{\hbar}{k_B}[6\pi^2 V^{1/2} n]^{1/3} f(\nu) \sqrt{\frac{B_S}{M}}, \tag{3}$$

where $n$ is the number of atoms per unit cell, $M$ is the unit cell mass, and $f(\nu)$ is a function of the Poisson ratio $\nu$:

$$f(\nu) = \left\{ 3 \left[ 2 \left( \frac{2}{3} \cdot \frac{1+\nu}{1-2\nu} \right)^{3/2} + \left( \frac{1}{3} \cdot \frac{1+\nu}{1-\nu} \right)^{3/2} \right]^{-1} \right\}^{\frac{1}{3}}, \tag{4}$$

where $\nu$ can be obtained from Eq. 2 using AEL or set directly by the user. The vibrational contribution to the free energy, $F_{vib}$, is given by:

$$F_{vib}(\theta_D; T) = nk_B T \left[ \frac{9}{8} \frac{\theta_D}{T} + 3 \log\left( 1 - e^{-\theta_D/T} \right) - D\left( \frac{\theta_D}{T} \right) \right], \tag{5}$$

where $D(\theta_D/T)$ is the Debye integral:

$$D(\theta_D/T) = 3 \left( \frac{T}{\theta_D} \right)^3 \int_0^{\theta_D/T} \frac{x^3}{e^x - 1} dx. \tag{6}$$

The Gibbs free energy is obtained from:

$$G(V; p, T) = E_{DFT}(V) + F_{vib}(\theta_D(V); T) + pV. \tag{7}$$

The volume which minimizes $G(V; p, T)$ at a given pressure $p$ and temperature $T$ is the equilibrium volume $V_{eq}$, which is used to evaluate $\theta_D(V_{eq})$ and the Grüneisen parameter $\gamma$ as defined by:

$$\gamma = -\frac{\partial \log(\theta_D(V))}{\partial \log V}. \tag{8}$$

Finally, $\theta_D$ and $\gamma$ are used to calculate other thermal properties including $C_V$, $C_p$, $\alpha_V$, and $\kappa_L$ (Toher et al. 2014; Blanco et al. 2004).

## 2.7   AFLOW-APL: Phonons

The AFLOW Automatic Phonon Library (AFLOW-APL) (Curtarolo et al. 2012a) calculates the harmonic vibrational properties of a crystal using the finite displacement method. Computed properties include the phonon dispersion and density of states, vibrational entropy ($S_{vib}$), and the heat capacity (at constant volume, $C_V$) as a function of temperature. These features are determined through an analysis of the phonon modes, accessed through the Interatomic Force Constants (IFCs) (Maradudin et al. 1971). To first approximation, the harmonic (second-order)

IFC $C_{i,j;\alpha,\beta}$ is the negative of the force exerted in the $\alpha$ direction on the atom $i$ when the atom $j$ is displaced in the $\beta$ direction, with all other atoms maintaining their equilibrium position. To determine the forces, the atoms of the structure (supercell) are individually perturbed as illustrated in Fig. 4d. The forces are obtained with DFT from the derivative of the total energy using the Hellmann-Feynman theorem. Supercells are used to sufficiently capture/isolate the impact of the distortion on the structure; distortions on small cells create forces on all atoms as well as their periodic images.

Given an input structure, AFLOW creates the full set of distorted supercell structures for the calculation of the forces. To minimize the number of expensive DFT calculations (primary computational bottleneck), AFLOW-SYM (see Sect. 2.3) is employed to determine which distortions are symmetrically equivalent using the site symmetry. Only inequivalent distortions are applied and explicitly calculated. Symmetry is then used to appropriately construct the IFC matrix, from which the dynamical matrix is constructed and the phonon modes, energies, and group velocities are derived.

APL has been extended to include the calculation of quasi-harmonic (quasi-harmonic approximation APL, QHA-APL Nath et al. 2016, 2017) and anharmonic (Automatic Anharmonic Phonon Library, AAPL Plata et al. 2017) effects in order to obtain properties such as the heat capacity at constant pressure $C_p$, coefficient of volumetric thermal expansion $\alpha_V$, and lattice thermal conductivity $\kappa_L$.

QHA-APL performs harmonic APL calculations at multiple different volumes and extracts the Grüneisen parameter from the change of the phonon frequencies with respect to volume:

$$\gamma_i = -\frac{V}{\omega_i}\frac{\partial\omega_i}{\partial V}. \tag{9}$$

The Grüneisen parameter can be used in combination with harmonic properties such as $C_V$ to calculate $C_p$, $\alpha_V$ (Nath et al. 2016), and $\kappa_L$ (Nath et al. 2017).

AAPL obtains the third-order anharmonic IFCs by distorting two atoms in a supercell structure at a time as depicted in Fig. 4e and then calculating the change in forces on the other atoms (Plata et al. 2017). These IFCs are used to calculate the three-phonon scattering rates, and thus the scattering time and mean free displacement. These quantities are combined with the group velocities obtained from harmonic APL to solve the Boltzmann transport equation and calculate $\kappa_L$ with quantitative accuracy (Plata et al. 2017).

## 2.8    AFLOW: Visualization Tools

AFLOW leverages a panoply of visualization tools for materials data, including standard software such as `gnuplot`, `latex`, and `xmgrace` for plots of phonon dispersions, electronic band structures, electronic density of states, and convex hull visualization. These plots are served publicly through the AFLOW.org repository.
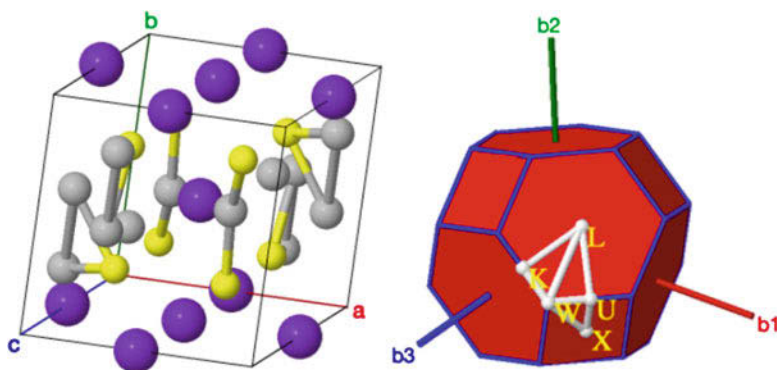
**Fig. 5 Side-by-size visualization of the crystal structure and Brillouin zone using Jmol.** The structure highlighted is $Ag_3KS_2$ (ICSD #73581): http://aflow.org/material.php?id=Ag6K2S4_ICSD_73581. The AFLOW Standard path of high-symmetry **k**-points is illustrated in the Brillouin zone (Setyawan and Curtarolo 2010)

To visualize crystal structures, AFLOW employs the Jmol software, which has incorporated substantial functionality for AFLOW-specific application. The JSmol branch of the software powers the online crystal structure visualizations in the AFLOW.org repository entry pages and AFLOW Prototype Library pages. With its recently added POSCAR reader, JSmol provides an assortment of capabilities ranging from different view perspectives, supercell expansions, and varying unit cell representations. A similar visualization application showing the AFLOW Standard high-symmetry paths in the Brillouin zone (Setyawan and Curtarolo 2010) is currently being incorporated, as illustrated in Fig. 5. Additionally, the Jmol desktop client offers a specialized macro (aflow) for visualization of alloy systems, which leverages the AFLUX Search-API.

## 3    AFLOW$\pi$: Minimalist High-Throughput

The AFLOW$\pi$ (Supka et al. 2017) framework has been originally implemented as a minimalist software to perform verification tasks (see Sect. 5.7) on data published on AFLOW.org. By design, AFLOW$\pi$ is easy to install and to extend to a variety of electronic structure codes (currently only the QUANTUM ESPRESSO (Giannozzi et al. 2009, 2017) DFT package is implemented). AFLOW$\pi$ builds on the versatility of Python, providing a module to prepare, run, and analyze large sets of first-principles calculations and includes tools for the automatic projection on pseudo-atomic orbitals (PAO; see Sect. 4) and the self-consistent calculation of Hubbard $U$ corrections within the Agapito, Curtarolo and Buongiorno Nardelli (ACBN0) approach (Agapito et al. 2015; Andrade et al. 2015). In addition, workflows for the calculation of elastic constants, diffusive transport coefficients, optical spectra, and phonon dispersions with DFT+$U$ (see Fig. 6a for assessing the
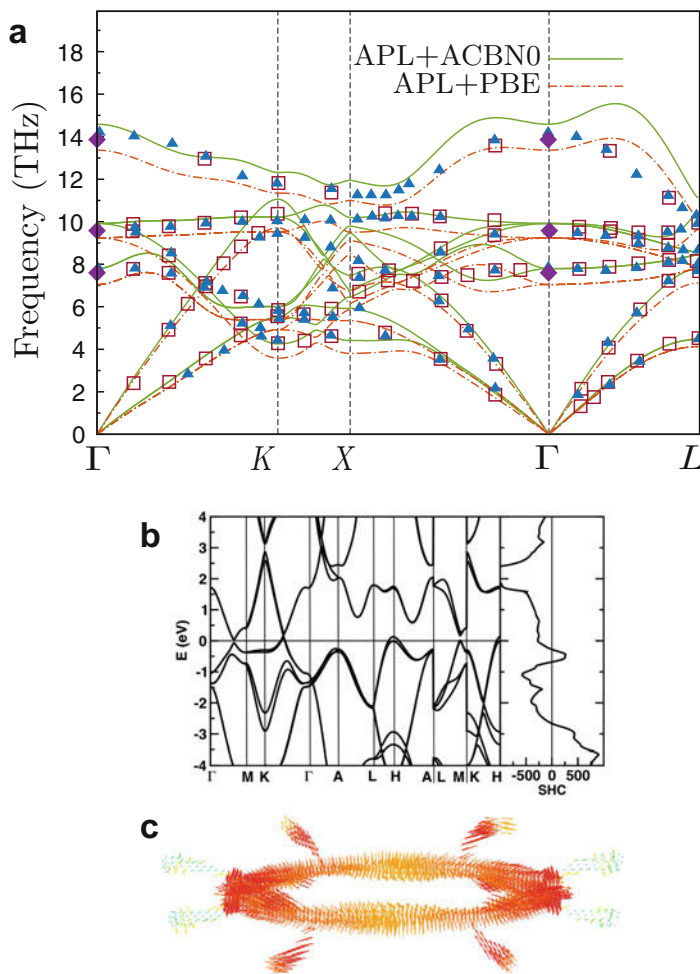
**Fig. 6 Vibrational spectrum calculated with AFLOW$\pi$ and AFLOW-APL (left) and electronic properties computed with PAOFLOW (right).** (**a**) Phonon dispersion of $CaF_2$ calculated with APL (Plata et al. 2017), using the ACBN0 method as implemented within AFLOW$\pi$ (green lines). The results obtained using PBE are shown by the broken orange lines for comparison. The blue triangles and red unfilled squares represent neutron scattering data from Schmalzl et al. (2003) and Elcombe and Pryor (1970), respectively, while the purple diamonds represent Raman and infrared data from Kaiser et al. (1962). (**b**) Electronic band structure, s̲pin H̲all c̲onductivity (SHC), and (**c**) spin texture of the nodal line and Weyl points in HfC, as calculated using the PAOFLOW utility

effect of the Hubbard $U$ corrections on the phonon dispersion calculated using APL) are included. When possible, AFLOW$\pi$ exploits the tight-binding hamiltonians as in D'Amico et al. (2016). Calculation results can be easily packaged and prepared for incorporation into the AFLOW.org data repository (see Sect. 5).

## 4      PAOFLOW: Fast Characterization

PAOFLOW (Buongiorno Nardelli et al. 2017) is a stand-alone tool to efficiently post-process standard DFT pseudo-potential plane-wave calculations to generate tight-binding (TB) Hamiltonians which faithfully reproduce the calculated electronic structure (eigenvalues and eigenvectors) with arbitrary precision in reciprocal space (Agapito et al. 2013, 2016a, b) (see Fig. 6b for PAOFLOW generated band structure for HfC (ICSD #169399, space group #187, AFLOW prototype: AB_hP2_187_d_a)). By exploiting the simplicity of the TB formalism and the efficiency of fast Fourier transforms, PAOFLOW interpolates the band structure and computes the matrix elements of the momentum operator, $\mathbf{p}_{m,n}$. These are used to improve the quality of integrated quantities such as the density of states (adaptive smearing), to compute electronic transport coefficients within the constant relaxation time approximation, and to compute the dielectric constants (D'Amico et al. 2016). In addition, the $\mathbf{p}_{m,n}$ matrix elements facilitate the calculation of the Berry curvature and related properties (anomalous Hall conductivity, spin Hall conductivity (see Fig. 6b), magnetic circular dichroism, spin circular dichroism; see spin texture of the nodal line and Weyl points in HfC shown in Fig. 6c). Starting from a well-interpolated band structure, it is also possible to compute topological invariants.

Because of the local representation of the electronic structure provided by the PAOFLOW software, surface-projected band structure and Landauer ballistic transport are also computable within PAOFLOW.

The software is implemented in Python, is portable and easy to install, and is parallel by design (on both CPUs and GPUs). PAOFLOW is also an integral part of the AFLOW$\pi$ framework.

## 5      AFLOW: Data Repository

The AFLOW data repository (Curtarolo et al. 2012b) contains the calculated properties for over 1.8 million materials entries, obtained using the AFLOW framework. These properties are available through the aflow.org web portal, which includes online search/sort and data analysis applications. The repository is programmatically accessible through the AFLOW Data REST-API (Taylor et al. 2014) and the AFLUX Search-API (Rose et al. 2017).

### 5.1      AFLOW: Web Portal

The AFLOW repository (Curtarolo et al. 2012b) is available online via the aflow.org web portal (Fig. 7a). It contains multiple online applications for data access, processing, and visualization, including the advanced "MendeLIB" search application at http://aflow.org/advanced.php which facilitates
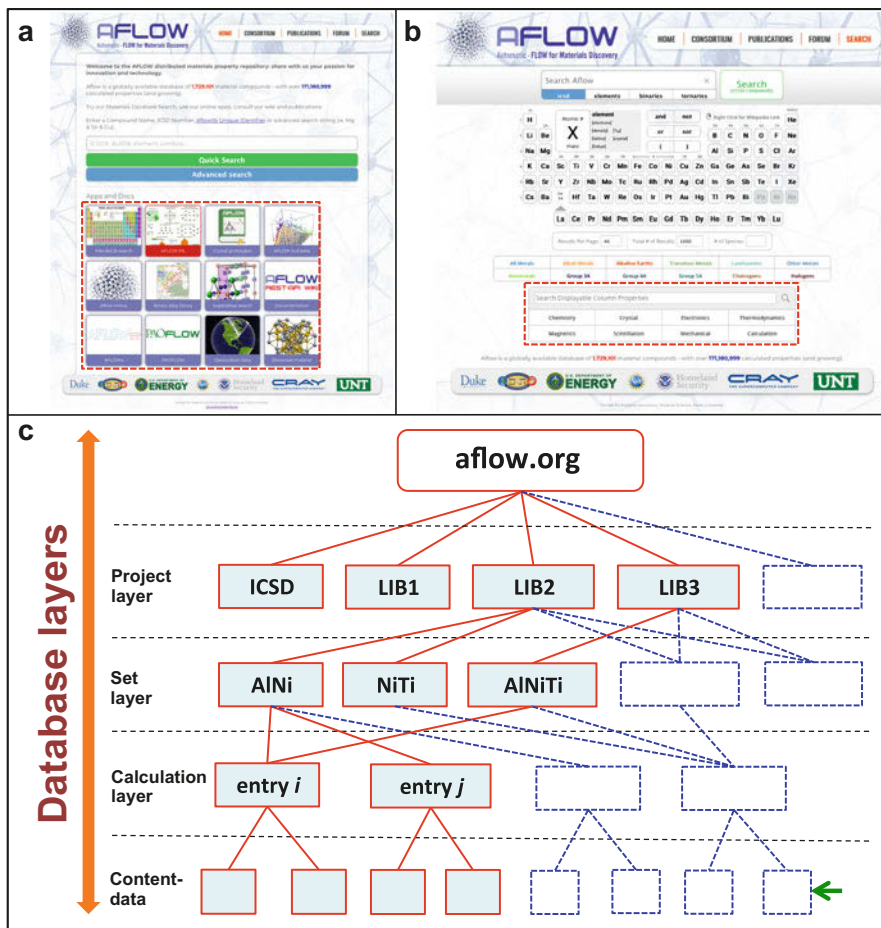
**Fig. 7 AFLOW web portal and data repository.** (**a**) Online applications and documentation are accessible via the "Apps and Docs" set of buttons surrounded by the dashed red rectangle. (**b**) The advanced search application can be used to search for specific compositions and also includes property search filters, as highlighted by the dashed red rectangle. (**c**) The AFLOW data repository is organized into project, set (i.e., alloy system), and calculation (i.e., materials entry) layers

searching for materials entries with filters for elemental composition and calculated properties (see Fig. 7b), the interactive convex hull application at `http://aflow.org/aflow-chull` (see Sect. 2.4), the online machine learning model at `http://aflow.org/aflow-ml` (see Sect. 5.2), and the AFLOW online tool at `http://aflow.org/aflow_online.html` which gives access to AFLOW crystal structure analysis and processing functions. These applications and more are accessible from the main `aflow.org` web page, via the "Apps and Docs" set of buttons highlighted by the dashed red rectangle in Fig. 7a.

## 5.2      AFLOW-ML: Online Machine Learning

The AFLOW machine learning (AFLOW-ML) online application provides a user interface to leverage machine learning models trained on AFLOW data. It accepts a standard structure file (POSCAR or QE) and outputs predictions for properties such as the band gap, elastic moduli, heat capacity, Debye temperature, vibrational free energy, and thermal expansion coefficient. Additionally, structures within the AFLOW repository can be imported via the sidebar. This application provides an accessible medium to retrieve machine learning predictions without the need to install a software library or machine learning package.

Currently, AFLOW-ML supports two different machine learning models. The first model, property-labeled materials fragments (Isayev et al. 2017), plmf, has been trained using data from the AFLOW repository and predicts properties such as the electronic band gap, specific heat capacities, and bulk/shear moduli. The second method is the molar fraction descriptor model (Legrain et al. 2017), mfd, which predicts vibrational properties such as vibrational free energy and entropy and is based only on the chemical composition of the material.

The AFLOW-ML API (Gossett et al. 2018) offers programmatic access to the AFLOW-ML online application and provides a simplified abstraction that facilitates leveraging powerful machine learning models. This distills the prediction process down to its essence: from a structure file, return a prediction. Using the API is a two-step process: first a structure file, in POSCAR 5 format (structure input for version 5 of VASP), is posted (i.e., uploaded) to the endpoint /<model>/prediction on the aflow.org server using standard HTTP libraries or dedicated programs such curl or wget, where <model> specifies the machine learning model to use in the prediction (current options: plmf and mfd). When a prediction is submitted, a JSON response object is returned that includes a task id. The results of the prediction are then retrieved from the /prediction/result/ endpoint on the aflow.org server by appending the task id to the end of the URL, i.e., /prediction/result/{id}/. This endpoint monitors the prediction task and responds with a JSON object that details its status. When complete, the endpoint responds with the results of the prediction, represented as a JSON object containing a key-value pair for each predicted property.

## 5.3      AFLOW: Database Organization

The AFLOW data repository (Curtarolo et al. 2012b) is organized into project, set, and calculation layers as illustrated in Fig. 7c. At the project layer, the calculations are divided into different catalogs based on the origin and composition of the entries (Taylor et al. 2014; Rose et al. 2017). Within each catalog, entries are grouped into sets based on shared lattice type or alloy system. The entries within each set contain the results of DFT calculated properties for particular structures.

The AFLOW-ICSD catalog contains the DFT calculated properties for over 57,000 experimentally observed materials listed in the Inorganic Crystal Structure Database (ICSD) (Bergerhoff et al. 1983; Karen and Hellenbrandt 2002). Internally, this catalog is organized by lattice type and then by individual materials entry. Since the materials in this catalog are already known to exist, the primary interest is in accurately calculating electronic structure and thermomechanical properties. Therefore, calculations for this catalog are generally performed using the Hubbard $U$ correction to the DFT exchange-correlation functional (Liechtenstein et al. 1995; Dudarev et al. 1998) where appropriate, using a set of standardized $U$ values (Calderon et al. 2015). Within this catalog, entries are grouped by Bravais lattice type into 14 sets: "BCC", "BCT", "CUB", "FCC", "HEX'", "MCL'", "MCLC'", "ORC'", "ORCC'", "ORCF'", "ORCI'", "RHL'", "TET'" and "TRI'". The name of each materials entry is generated using the format `<composition>_ICSD_<ICSD number>`.

The entries in the other catalogs, such as "LIB1," "LIB2," and "LIB3", are generated by decorating crystal structure prototypes to predict new hypothetical compounds and contain unary, binary, and ternary materials, respectively. Additional catalogs, "LIB4," "LIB5," and "LIB6," are currently being generated for quaternary, quinary, and hexenary materials. Within each catalog, the entries are grouped by element and exchange-correlation functional in the case of "LIB1" and by alloy system in the cases of "LIB2" and "LIB3." "LIB1" contains a total of 3068 entries, while "LIB2" currently has 329,192 entries and "LIB3" has over 1.4 million. Within each alloy system, the individual materials entries are named according to the relevant crystal prototype. For these catalogs, the emphasis is on the discovery of new thermodynamically stable or metastable materials and on their use to generate the thermodynamic density of states for the prediction of the formation of disordered materials such as metallic glasses (Perim et al. 2016) or high-entropy alloys (Lederer et al. 2018). Therefore, calculations in these catalogs are performed using the GGA-PBE exchange-correlation functional (Perdew et al. 1996) without Hubbard $U$ corrections (Calderon et al. 2015) so as to produce consistent energy differences, enabling the calculation of accurate formation enthalpies.

## 5.4 AFLOW: Database Properties

Materials properties within the AFLOW repository (Curtarolo et al. 2012b) are indexed as keyword-value pairs which are programmatically accessible via the AFLOW Data REST-API (Taylor et al. 2014) and programmatically searchable via the AFLUX Search-API (Rose et al. 2017). Search filters for these properties are also available in the advanced search application of the aflow.org web portal as highlighted by the dashed red rectangle in Fig. 7b, where they are grouped into chemistry (e.g., chemical species, stoichiometry), crystal (e.g., space group, Bravais lattice type), electronic (e.g., band gaps), thermodynamic (energetic and thermal properties, e.g., formation enthalpies and Debye temperatures), magnetic, scintillation, mechanical (elastic moduli and pressure-related properties), and calculation (e.g., **k**-point mesh, AFLOW version) parameters. In total there are in excess of 170
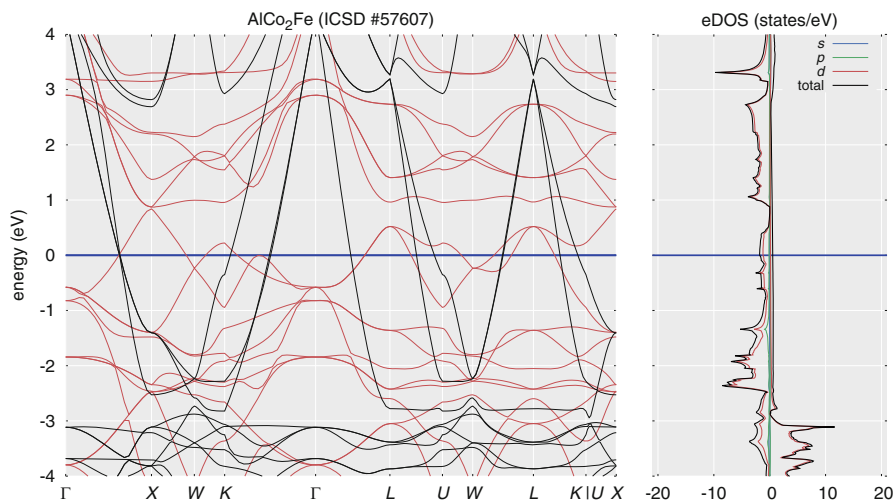
**Fig. 8 Example band structure and density of states images automatically generated and served through the AFLOW.org data repository.** The structure highlighted is AlCo$_2$Fe (ICSD #57607): http://aflow.org/material.php?id=Al1Co2Fe1_ICSD_57607. The results of the spin-polarized calculation are differentiated by color on the band structure plot (black/red for majority/minority spin) and sign on the density of states plot (positive/negative for majority/minority spin). The band structure is calculated following the AFLOW Standard path of high-symmetry **k**-points (Setyawan and Curtarolo 2010)

million individual materials properties indexed in the AFLOW database ($\sim$100 per materials entry). Lists of the keywords corresponding to the materials properties are provided in Taylor et al. (2014), Rose et al. (2017), and Toher et al. (2017).

Systems for which the "STATIC" and "BANDS" calculations have been performed are supplemented with automatically generated images of the density of states, projected density of states, and band structure. Both low (PNG)- and high (EPS)-quality variants of the images are available for download. An example band structure and density of states image is displayed in Fig. 8.

## 5.5   AFLOW: Data REST-API

The full data set generated by the high-throughput AFLOW process (Curtarolo et al. 2012b) is backed by a disk store of (at this time) over 12 TB of input criteria, calculated results, and derivative output. The backing store is exposed via the AFLOW Data REST-API (Curtarolo et al. 2012b) in a hierarchical organization. This direct exposure of our results not only grants the end user a high degree of utility via direct access, but, more importantly, guarantees data provenance that promotes reproducibility. The hierarchy of the AFLOW Data REST-API categorizes this abundance of information into meaningful high-level classifications allowing

for exploration of self-similar materials that are related by stoichiometric and/or crystallographic properties. Once a selection of materials has been determined, the full range of available properties and procedural data are retrievable.

The organizational hierarchy of both the underlying data store and the REST-API is project dependent, as described in Sect. 5.3. Each project is equivalent to one of the catalogs listed in Sect. 5.3 and in the REST-API is denoted by the project layers "ICSD_WEB," "LIB1_RAW," "LIB2_RAW," and "LIB3_RAW." Each project layer contains multiple set layers, which correspond to specific alloy systems in the case of "LIB1_RAW," "LIB2_RAW," and "LIB3_RAW." For instance, `http://aflowlib.duke.edu/AFLOWDATA/LIB2_RAW/` exposes the set layer for binary entries, where each set corresponds to different binary alloy systems, allowing for pairwise atomic species examination. Within each set is the entry layer, consisting of decorated structural prototypes which provide stoichiometric and structural variation for each alloy system. Each entry contains the calculated results for a particular structure and composition, organized as keyword-value pairs. The calculated values of thermodynamical, mechanical, electronic, magnetic, chemical, and crystallographic properties can be directly accessed by querying a Uniform Resource Identifier (URI) of the form `<server>/<project>/<set>/<entry>/?<keyword>`, where `<server>` is `http://aflowlib.duke.edu/AFLOWDATA`, `<project>` is the appropriate project layer, `<set>` is the alloy system, `<entry>` is the structural prototype, and `<keyword>` corresponds to the materials property of interest. A full description of the REST-API keywords is provided in Taylor et al. (2014), along with additions in the appendices of Rose et al. (2017) and Toher et al. (2017).

The ability to explore related entries predicated on a multitude of properties leads directly to novel materials discovery and use. The AFLOW Data REST-API disseminates our methods and results, without restriction, to a global research audience in order to promote scientific and engineering advancement.

## 5.6    AFLUX: Search-API

The Automatic Flow of LUX or AFLUX Search-API (Rose et al. 2017) is a human usable remote data search API. LUX is designed to be a domain agnostic solution to the outstanding problem of programmatically searching remote data that typically is either exposed via a capriciously limited utility or requires a-priori knowledge of the internal organization of the remote repository. The LUX query concept flattens the exposed data, while simultaneously providing arbitrarily complex query capability, allowing an end user full freedom in constraining the requested data. LUX is designed to operate in the nearly ubiquitous web URI context while minimizing any potentially conflicting interactions with existing URI functionality.

AFLUX is the domain-specific implementation of LUX and is available at the `<AFLUX-URI>`: `http://aflowlib.duke.edu/search/API/?`. At this time, the AFLUX API freely exposes over 180 million keyword-value properties without any requirements or restrictions on the end user. Specific properties and

compositions can be searched for by appending the appropriate keywords to the `<AFLUX-URI>`. Search results can be restricted by including specific values or value ranges in parentheses after the appropriate keyword. For example, a search can be restricted to entries that contain both of the elements Na and Cl, and have a calculated electronic band gap in excess of 1.0eV, by including the search parameters `species(Na,Cl),Egap(1.0*)` in the query part of the URI. In LUX, "," corresponds to the logical AND operator, and "*" is the `loose` operator which extends the search to entries in a specified value range. If no parameters are provided for a particular keyword, then the values of that property are returned for all entries which satisfy the remaining search criteria. A full list of all LUX logical operators can be found in Rose et al. (2017), along with descriptions of their functionality and appropriate usage.

In addition to materials properties keywords, LUX also accepts directives, which behave as pseudo property keywords. They are used to provide additional information on LUX usage and control the format and quantity of the returned data. Note that any directives included in a search query must come after all of the materials properties keywords. In particular, the `schema` directive can be used to retrieve the most current and canonical list of keywords using the AFLUX summons: `<AFLUX-URI>schema,format(json)`.

## 5.7   AFLOW: Data Quality Control

Data quality control, including validation of methodologies and verification of calculated data, is vital when constructing large databases such as the AFLOW repository (Curtarolo et al. 2012b) in order to guarantee the reliability of the results. Methodological validation involves quantifying the accuracy of calculation models with respect to experiment, while data verification includes checking the robustness of calculation parameters and the satisfaction of convergence criteria.

Physical models incorporated into the AFLOW framework are validated by comparison to benchmark sets of experimental data. This helps determine the predictive accuracy of the methods for real materials, as well as the regimes in which they are reliable. For example, the AEL and AGL modules were validated by comparison to a benchmark set of ∼75 experimentally well-characterized compounds of various structural types (Toher et al. 2014, 2017), and the accuracy was quantified by the Pearson and Spearman correlations and the root-mean-square deviations. Similar validation analyses were performed for the QHA-APL (Nath et al. 2016, 2017) and AAPL methods (Plata et al. 2017), as well as the property labeled materials fragments machine learning model (Isayev et al. 2017).

The AFLOW-POCC methodology has been validated by comparing the band gap as a function of composition for $ZnS_{1-x}Se_x$ and $Mg_xZn_{1-x}O$, and the magnetic moment per atom as a function of composition for $Fe_{1-x}Cu_x$, to experimental values (Yang et al. 2016).

The ACBN0 functional (Agapito et al. 2015), implemented in AFLOW$\pi$ (see Sect. 3) and PAOFLOW (see Sect. 4), has been validated by comparing the lattice

parameters, bulk moduli, electronic band gaps, phonon modes, high-frequency dielectric constants, and Born effective changes it produces to the experimentally measured values for the Zn and Cd chalcogenides (Gopal et al. 2015).

The convergence of both the charge density optimization and the ionic structural relaxation are automatically verified for all AFLOW calculations prior to incorporation into the data repository. This includes, for example, checking that the charge density has converged in accordance with the AFLOW Standard settings (Calderon et al. 2015) and verifying the relaxation of the cell size and shape by ensuring that all elements of the stress tensor are less than 10kB. The convergence level for any individual calculation can be verified by querying appropriate keywords for the stress tensor: `stress_tensor`, Pulay stress: `Pulay_stress`, residual external pressure on the relaxed cell: `pressure_residual`, and the $\delta E$ value for the final electronic convergence step: `delta_electronic_energy_convergence`, using the AFLOW Data REST-API (Taylor et al. 2014) or the AFLUX Search-API (Rose et al. 2017). Initial calculation parameters can similarly be obtained using the appropriate keywords, such as the k-point grid, `kpoints`, or the electronic energy convergence threshold, `delta_electronic_energy_threshold`.

## Conclusion

The AFLOW Fleet for computational materials design automates first-principles calculations of materials properties. AFLOW incorporates a wide range of different modules, including applications for symmetry and thermodynamic stability analysis, generation of ordered and disordered materials structures, and calculation of thermomechanical properties, in a single integrated framework. AFLOW$\pi$ is a versatile minimalist framework that includes tools for projection onto pseudo-atomic orbitals (PAO) and the self-consistent calculation of Hubbard $U$ corrections using ACBN0. PAOFLOW generates tight-binding Hamiltonians which reproduce the electronic structure calculated using first-principles methods, facilitating the rapid calculation of electronic and magnetic properties such as transport coefficients and the Berry curvature. All results are stored in, and disseminated through, the AFLOW data repository, which is available online at aflow.org and is programmatically accessible via the AFLOW Data REST-API and the AFLUX Search-API.

# References

Agapito LA, Ferretti A, Calzolari A, Curtarolo S, Buongiorno Nardelli M (2013) Effective and accurate representation of extended Bloch states on finite Hilbert spaces. Phys Rev B 88:165127

Agapito LA, Curtarolo S, Buongiorno Nardelli M (2015) Reformulation of DFT + $U$ as a pseudohybrid Hubbard density functional for accelerated materials discovery. Phys Rev X 5:011006

Agapito LA, Fornari M, Ceresoli D, Ferretti A, Curtarolo S, Buongiorno Nardelli M (2016a) Accurate tight-binding Hamiltonians for two-dimensional and layered materials. Phys Rev B 93:125137

Agapito LA, Ismail-Beigi S, Curtarolo S, Fornari M, Buongiorno Nardelli M (2016b) Accurate tight-binding Hamiltonian matrices from *ab initio* calculations: minimal basis sets. Phys Rev B 93:035104

Andrade X, Strubbe D, De Giovannini U, Larsen AH, Oliveira MJT, Alberdi-Rodriguez J, Varas A, Theophilou I, Helbig N, Verstraete MJ, Stella L, Nogueira F, Aspuru-Guzik A, Castro A, Marques MAL, Rubio A (2015) Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. Phys Chem Chem Phys 17:31371–31396

Bahn SR, Jacobsen KW (2002) An object-oriented scripting interface to a legacy electronic structure code. Comput Sci Eng 4:56–66

Barber CB, Dobkin DP, Huhdanpaa H (1996) The quickhull algorithm for convex hulls. ACM Trans Math Soft 22:469–483

Barzilai S, Toher C, Curtarolo S, Levy O (2016) Evaluation of the tantalum-titanium phase diagram from *ab-initio* calculations. Acta Mater 120:255–263

Barzilai S, Toher C, Curtarolo S, Levy O (2017a) The effect of lattice stability determination on the computational phase diagrams of intermetallic alloys. J Alloys Compd 728:314–321

Barzilai S, Toher C, Curtarolo S, Levy O (2017b) Molybdenum-titanium phase diagram evaluated from *ab initio* calculations. Phys Rev Mater 1:023604

Bergerhoff G, Hundt R, Sievers R, Brown ID (1983) The inorganic crystal structure data base. J Chem Inf Comput Sci 23:66–69

Blanco MA, Martín-Pendás A, Francisco E, Recio JM, Franco R (1996) Thermodynamical properties of solids from microscopic theory: applications to $MgF_2$ and $Al_2O_3$. J Mol Struct Theochem 368:245–255

Blanco MA, Francisco E, Luaña V (2004) GIBBS: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model. Comput Phys Commun 158:57–72

Blöchl PE (1994) Projector augmented-wave method. Phys Rev B 50:17953–17979

Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M (2009) Ab initio molecular simulations with numeric atom-centered orbitals. Comput Phys Commun 180:2175–2196

Calderon CE, Plata JJ, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl MJ, Hart GLW, Buongiorno Nardelli M, Curtarolo S (2015) The AFLOW standard for high-throughput materials science calculations. Comput Mater Sci 108(Part A):233–238

Curtarolo S, Setyawan W, Hart GLW, Jahnátek M, Chepulskii RV, Taylor RH, Wang S, Xue J, Yang K, Levy O, Mehl MJ, Stokes HT, Demchenko DO, Morgan D (2012a) AFLOW: an automatic framework for high-throughput materials discovery. Comput Mater Sci 58:218–226

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, Nelson LJ, Hart GLW, Sanvito S, Buongiorno Nardelli M, Mingo N, Levy O (2012b) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput *ab initio* calculations. Comput Mater Sci 58: 227–235

Curtarolo S, Hart GLW, Buongiorno Nardelli M, Mingo N, Sanvito S, Levy O (2013) The high-throughput highway to computational materials design. Nat Mater 12:191–201

D'Amico P, Agapito LA, Catellani A, Ruini A, Curtarolo S, Fornari M, Buongiorno Nardelli M, Calzolari A (2016) Accurate *ab initio* tight-binding Hamiltonians: effective tools for electronic transport and optical spectroscopy from first principles. Phys Rev B 94:165166

de Jong M, Chen W, Angsten T, Jain A, Notestine R, Gamst A, Sluiter M, Ande CK, van der Zwaag S, Plata JJ, Toher C, Curtarolo S, Ceder G, Persson KA, Asta MD (2015) Charting the complete elastic properties of inorganic crystalline compounds. Sci Data 2:150009

Dudarev SL, Botton GA, Savrasov SY, Humphreys CJ, Sutton AP (1998) Electron-energy-loss spectra and the structural stability of Nickel oxide: an LSDA+$U$ study. Phys Rev B 57:1505–1509

Elcombe MM, Pryor AW (1970) The lattice dynamics of calcium fluoride. J Phys C Solid State Phys 3:492

Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. J Phys Condens Matter 21:395502

Giannozzi P, Andreussi O, Brumme T, Bunau O, Buongiorno Nardelli M, Calandra M, Car R, Cavazzoni C, Ceresoli D, Cococcioni M, Colonna N, Carnimeo I, Dal Corso A, de Gironcoli S, Delugas P, DiStasio Jr RA, Ferretti A, Floris A, Fratesi G, Fugallo G, Gebauer R, Gerstmann U, Giustino F, Gorni T, Jia J, Kawamura M, Ko H-Y, Kokalj A, Küçükbenli E, Lazzeri M, Marsili M, Marzari N, Mauri F, Nguyen NL, Nguyen H-V, Otero-de-la-Roza A, Paulatto L, Poncé S, Rocca D, Sabatini R, Santra B, Schlipf M, Seitsonen AP, Smogunov A, Timrov I, Thonhauser T, Umari P, Vast N, Wu X, Baroni S (2017) Advanced capabilities for materials modelling with quantum ESPRESSO. J Phys Condens Matter 29:465901

Gonze X, Beuken JM, Caracas R, Detraux F, Fuchs M, Rignanese GM, Sindic L, Verstraete M, Zerah G, Jollet F, Torrent M, Roy A, Mikami M, Ghosez P, Raty JY, Allan DC (2002) First-principles computation of material properties: the ABINIT software project. Comput Mater Sci 25:478–492

Gopal P, Fornari M, Curtarolo S, Agapito LA, Liyanage LSI, Buongiorno Nardelli M (2015) Improved predictions of the physical properties of Zn- and Cd-based wide band-gap semiconductors: a validation of the ACBN0 functional. Phys Rev B 91:245202

Gossett E, Toher C, Oses C, Isayev O, Legrain F, Rose F, Zurek E, Carrete J, Mingo N, Tropsha A, Curtarolo S (2018) AFLOW-ML: a RESTful API for machine-learning predictions of materials properties. Comput Mater Sci 152:134–145

Hart GLW, Forcade RW (2008) Algorithm for generating derivative structures. Phys Rev B 77:224115

Hicks D, Oses C, Gossett E, Gomez G, Taylor RH, Toher C, Mehl MJ, Levy O, Curtarolo S (2018) AFLOW-SYM: platform for the complete, automatic and self-consistent symmetry analysis of crystals. Acta Cryst A 74:184–203

Isayev O, Fourches D, Muratov EN, Oses C, Rasch K, Tropsha A, Curtarolo S (2015) Materials cartography: representing and mining materials space using structural and electronic fingerprints. Chem Mater 27:735–743

Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A (2017) Universal fragment descriptors for predicting electronic properties of inorganic crystals. Nat Commun 8:15679

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:011002

Kaiser W, Spitzer WG, Kaiser RH, Howarth LE (1962) Infrared Properties of $CaF_2$, $SrF_2$, and $BaF_2$. Phys Rev 127:1950–1954

Karen VL, Hellenbrandt M (2002) Inorganic crystal structure database: new developments. Acta Cryst A58:c367

Kresse G, Furthmüller J (1996) Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. Phys Rev B 54:11169–11186

Kresse G, Hafner J (1993) *Ab initio* molecular dynamics for liquid metals. Phys Rev B 47:558–561

Landis DD, Hummelshøj JS, Nestorov S, Greeley J, Dułak M, Bligaard T, Nørskov JK, Jacobsen KW (2012) The computational materials repository. Comput Sci Eng 14:51–57

Lederer Y, Toher C, Vecchio KS, Curtarolo S (2018) The search for high entropy alloys: a high-throughput *ab-initio* approach. Acta Mater 159:354–383

Legrain F, Carrete J, van Roekeghem A, Curtarolo S, Mingo N (2017) How chemical composition alone can predict vibrational free energies and entropies of solids. Chem Mater 29:6220–6227

Levy O, Hart GLW, Curtarolo S (2010a) Uncovering compounds by synergy of cluster expansion and high-throughput methods. J Am Chem Soc 132:4830–4833

Levy O, Hart GLW, Curtarolo S (2010b) Hafnium binary alloys from experiments and first principles. Acta Mater 58:2887–2897

Levy O, Chepulskii RV, Hart GLW, Curtarolo S (2010c) The new face of Rhodium alloys: revealing ordered structures from first principles. J Am Chem Soc 132:833–837

Liechtenstein AI, Anisimov VI, Zaanen J (1995) Density-functional theory and strong interactions: orbital ordering in Mott-Hubbard insulators. Phys Rev B 52:R5467–R5470

Maradudin AA, Montroll EW, Weiss GH, Ipatova IP (1971) Theory of lattice dynamics in the harmonic approximation. Academic Press, New York

Mehl MJ, Hicks D, Toher C, Levy O, Hanson RM, Hart GLW, Curtarolo S (2017) The AFLOW library of crystallographic prototypes: part 1. Comput Mater Sci 136:S1–S828

Buongiorno Nardelli M, Cerasoli FT, Costa M, Curtarolo S, De Gennaro R, Fornari M, Liyanage L, Supka AR, Wang H (2017) PAOFLOW: a utility to construct and operate on ab initio Hamiltonians from the projections of electronic wavefunctions on atomic orbital bases, including characterization of topological materials. Comput Mater Sci 143:462–472

Nath P, Plata JJ, Usanmaz D, Al Rahal Al Orabi R, Fornari M, Buongiorno Nardelli M, Toher C, Curtarolo S (2016) High-throughput prediction of finite-temperature properties using the quasi-harmonic approximation. Comput Mater Sci 125:82–91

Nath P, Plata JJ, Usanmaz D, Toher C, Fornari M, Buongiorno Nardelli M, Curtarolo S (2017) High throughput combinatorial method for fast and robust prediction of lattice thermal conductivity. Scr Mater 129:88–93

Nyshadham C, Oses C, Hansen JE, Takeuchi I, Curtarolo S, Hart GLW (2017) A computational high-throughput search for new ternary superalloys. Acta Mater 122:438–447

Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier VL, Persson KA, Ceder G (2013) Python Materials Genomics (pymatgen): a robust, open-source Python library for materials analysis. Comput Mater Sci 68:314–319

Oses C, Gossett E, Hicks D, Rose F, Mehl MJ, Perim E, Takeuchi I, Sanvito S, Scheffler M, Lederer Y, Levy O, Toher C, Curtarolo S (2018) AFLOW-CHULL: Cloud-oriented platform for automonous phase stability analysis. J Chem Inf Model 58:2477–2490 https://doi.org/10.1021/acs.jcim.8b00393

Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865–3868

Perim E, Lee D, Liu Y, Toher C, Gong P, Li Y, Simmons WN, Levy O, Vlassak JJ, Schroers J, Curtarolo S (2016) Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases. Nat Commun 7:12315

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci 111:218–230

Plata JJ, Nath P, Usanmaz D, Carrete J, Toher C, de Jong M, Asta MD, Fornari M, Buongiorno Nardelli M, Curtarolo S (2017) An efficient and accurate framework for calculating lattice thermal conductivity of solids: AFLOW-AAPL automatic anharmonic phonon library. NPJ Comput Mater 3:45

Poirier J-P (2000) Introduction to the physics of the earth's interior, 2nd edn. Cambridge University Press, Cambridge

Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc 114:10024–10035

Rose F, Toher C, Gossett E, Oses C, Buongiorno Nardelli M, Fornari M, Curtarolo S (2017) AFLUX: the LUX materials search API for the AFLOW data repositories. Comput Mater Sci 137:362–370

Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). JOM 65:1501–1509

Sanvito S, Oses C, Xue J, Tiwari A, Zic M, Archer T, Tozman P, Venkatesan M, Coey JMD, Curtarolo S (2017) Accelerated discovery of new magnets in the Heusler alloy family. Sci Adv 3:e1602241

Scheffler M, Draxl C (2014) The NoMaD Repository. http://nomad-repository.eu

Schmalzl K, Strauch D, Schober H (2003) Lattice-dynamical and ground-state properties of $CaF_2$ studied by inelastic neutron scattering and density-functional methods. Phys Rev B 68:144301

Setyawan W, Curtarolo S (2010) High-throughput electronic band structure calculations: challenges and tools. Comput Mater Sci 49:299–312

Supka AR, Lyons TE, Liyanage LSI, D'Amico P, Al Rahal Al Orabi R, Mahatara S, Gopal P, Toher C, Ceresoli D, Calzolari A, Curtarolo S, Buongiorno Nardelli M, Fornari M (2017) AFLOW$\pi$: a minimalist approach to high-throughput *ab initio* calculations including the generation of tight-binding Hamiltonians. Comput Mater Sci 136:76–84

Taylor RH, Rose F, Toher C, Levy O, Yang K, Buongiorno Nardelli M, Curtarolo S (2014) A RESTful API for exchanging materials data in the AFLOWLIB.org consortium. Comput Mater Sci 93:178–192

Toher C, Plata JJ, Levy O, de Jong M, Asta MD, Buongiorno Nardelli M, Curtarolo S (2014) High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasiharmonic Debye model. Phys Rev B 90:174107

Toher C, Oses C, Plata JJ, Hicks D, Rose F, Levy O, de Jong M, Asta MD, Fornari M, Buongiorno Nardelli M, Curtarolo S (2017) Combining the AFLOW GIBBS and elastic libraries to efficiently and robustly screen thermomechanical properties of solids. Phys Rev Mater 1:015401

Yang K, Oses C, Curtarolo S (2016) Modeling off-stoichiometry materials with a high-throughput *ab-initio* approach. Chem Mater 28:6484–6492

# Open-Science Platform for Computational Materials Science: AiiDA and the Materials Cloud

# 83

Giovanni Pizzi

## Contents

G. Pizzi (✉)
Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: giovanni.pizzi@epfl.ch

**Abstract**

We discuss here our vision for an Open-Science platform for computational materials science. Such a platform needs to rely on three pillars, consisting of (1) open data generation tools (including the simulation codes, the scientific workflows, and the infrastructure for automation and provenance tracking), (2) an open integration platform where these tools interact in an easily accessible way and computations are coordinated by automated workflows, and (3) support for seamless code and data sharing through portals that are FAIR-compliant and compatible with data management plans. As a practical implementation, we show how such a platform can be achieved in a few examples and focusing in particular on the combination of the AiiDA infrastructure and the Materials Cloud web portal.

# 1    Introduction

In the field of atomistic materials science, computer simulations have become a key ingredient in materials design. The availability of accurate codes based on density-functional theory (DFT) and beyond-DFT methods and the ever-increasing speed of supercomputers make computational materials science more accessible. Indeed, computations have developed into an essential component to complement experiments in the study and optimization of materials properties. The current relevance of these techniques is also demonstrated by the fact that 12 out of the top-100 most-cited papers in the whole scientific literature are about DFT-related methods (Van Noorden et al. 2014). As a consequence, many groups have started computing materials properties for large sets of known and unknown materials, often starting from databases of crystal structures like the ICSD (Belsky et al. 2002), the COD (Gražulis et al. 2012), and the Pauling File (Villars et al. 2004). Many of these computed materials properties are available online; some of these portals are also described in this handbook and include the Materials Project (Jain et al. 2011), AFLOWlib (Curtarolo et al. 2012), OQMD (Saal et al. 2013), Nomad (The Nomad Repository 2018), and the CMR (Landis et al. 2012).

Furthermore, large sets of computations are nowadays being used as training data to predict materials properties more efficiently using machine-learning techniques (Ramakrishnan et al. 2014; Dragoni et al. 2018). In general, however, it is essential to have enough information on the data, including how it was generated (i.e., its provenance), which physical and numerical parameters were used, and in general be able to reproduce the results to validate them. This is even more important for machine-learning, where accurate predictions are possible only if the quality of the data used to train the algorithms is known and consistent across datapoints.

In addition to this, the availability of open data is beneficial to boost research and discovery because datasets can be repurposed for new studies and analyses not considered by the original authors. In principle, recording data with its full provenance and sharing it should be much easier for computer simulations with respect to, e.g., experiments. However, in practice there are a number of aspects that

hinder automatic computation of materials properties, calculation reproducibility, dissemination of data provenance, and sharing of open research data. In this chapter, we discuss our vision of a platform for Open Science that can lower these barriers, and we review the challenges that need to be addressed. We also present the two software infrastructures AiiDA (Pizzi et al. 2016) and Materials Cloud (Materials Cloud 2017) and show examples of how their combination makes it possible to create a fruitful Open Science ecosystem.

## 1.1    The Pillars of an Open-Science Platform

Before starting, we want to define the term Open-Science platform (OSP) and how it is used in this chapter. With Open Science, we refer to a combination of open tools and data that make it possible to run simulations and then share and reuse the results without barriers, with the aim to accelerate scientific discovery. While open data is definitely an essential ingredient of an OSP, we believe that the platform must have a strong focus on the tools to generate and share the data. Moreover, it should be composed by modular components, so as to cover a multitude of use-cases and to encourage researchers to use and expand it with contributions.

We think that an OSP should be based around the three following pillars: (1) open data generation tools, including open simulation codes, an open architecture to manage simulations, and open workflows to steer them; (2) an integration platform that makes these tools accessible and available in the form of automated solutions, not only to experienced computational researchers but also to experimentalists, students, or the industry; and (3) support for seamless data sharing through portals that make data not only findable (e.g., via DOIs) and openly available, but also interoperable and reusable, encouraging the use of open data and code licenses.

Moreover, in our vision an OSP should also include the availability of open libraries of curated input data, often needed for simulations (like crystal structures or pseudopotentials) and that can enable the creation of automated workflows, as well as of open learning and educational resources to ease the introduction of young researchers to the field.

## 2    Open Science Pillar 1: Open Data Generation

## 2.1    Open Simulation Tools

The first requirement to be able to generate data within an OSP is the availability of open simulation codes. In the field of materials science (and limiting to atomistic simulations using density-functional theory only), a number of open codes are available thanks to the developments that have happened in the past few decades. These include Quantum ESPRESSO (Giannozzi et al. 2017), SIESTA (Soler et al. 2002), YAMBO (Marini et al. 2009), FLEUR (Blügel and Bihlmayer 2006), CP2K (Hutter et al. 2014), and ABINIT (Gonze et al. 2016) just to mention a few.

This open licensing model makes the codes accessible to everybody. Moreover, it becomes possible to build simulation services on top of them that can also be beneficial to the code themselves, as these services can become financing channels for the code development.

There are also other codes that are widespread in the community and have commercial licenses, like, for instance, VASP (Kresse and Furthmüller 1996) or CASTEP (Clark et al. 2005). These might have benefits in term of additional features implemented, speed, robustness, more widespread adoption, or stronger user support (the latter is often very valued outside academia in industrial research and development environments). The challenge for an OSP becomes then to be able to integrate also these non-open tools while abiding by their license terms. Models can be devised that are beneficial both for the commercial codes and for the platform. An example could be to provide open interfaces and plugins for the codes to enable or facilitate their integration in the OSP, while keeping a commercial license for the codes. The latter can also benefit from this model because having interfaces ready can facilitate the code adoption by OSP users.

## 2.2    The ADES Model and the Implementation in AiiDA

In recent years, many research projects used a high-throughput computing (HTC) approach to scan hundreds of thousands of different systems and identify those with optimal materials properties. For this kind of projects, it is unrealistic to run all the simulations manually and even more to control the sequence of calculations needed to compute a given materials property. Tools are hence needed to help manage and store simulations, search through them, and at the same time steer their execution when calculation dependencies exist. Our experience showed that these tools can easily grow in complexity if they need to be reusable and modular. Therefore, there is a need to collect and develop them in an organized architecture.

In Pizzi et al. (2016), a model for such a computational science architecture to manage calculations and workflows has been discussed, based on the four ADES pillars of automation, data, environment, and sharing. These are at the foundation of AiiDA, a python platform introduced in the same paper. Here, we briefly describe the ADES model and how AiiDA implements its four pillars, and later we discuss why an ADES-compliant architecture is essential within an OSP.

The first ADES pillar, *Automation*, involves all those software components that aim at solving the issue of managing large numbers of HTC runs on supercomputers. AiiDA, in order to be independent of the supercomputer details, implements plugin interfaces to control the connection, transfer files, and execute commands (e.g., via SSH) or to interact with job schedulers. These plugins are used by a daemon that runs in the background and is responsible for creating new calculation inputs and uploading them to the supercomputers, submitting new simulations and managing their lifecycle on a job scheduler, and retrieving and parsing results when they finish.

The second ADES pillar, *Data*, is then needed to store and preserve all the generated data in a reproducible and searchable way. AiiDA uses a provenance

model based on directed acyclic graphs to keep track of all inputs and outputs and of the logical relationships between different calculations. The provenance is tracked automatically by AiiDA (see also Sect. 2.3) and can be browsed at any time to understand how data were generated or calculations were run. An example of a provenance graph as tracked by AiiDA is shown in Fig. 1. Moreover, it is also important to have the possibility to analyze results efficiently. To this aim, AiiDA stores the provenance graph in a database, whose schema is optimized to ensure that typical queries run fast.

The third ADES pillar, *Environment*, focuses on two aspects. The first consists in ensuring that AiiDA is simple to use and provides users with an effective research environment to help and facilitate them in their work. This happens thanks to user APIs and command-line tools that simplify the execution of typical tasks (management of computational resources and codes, inspection of calculation inputs and results, data analytics) and by extensive documentation of these features. Moreover, AiiDA provides a transparent access to the database that does not require knowledge of SQL or similar database languages, to perform queries and data analytics. This is implemented both via the use of object-relational mapping (ORM) classes to access directly the data stored in the database and via the QueryBuilder class to enable users to perform generic queries on the AiiDA graph by using standard python syntax, independent of the database backend. The other defining aspect of the environment pillar is the possibility to implement and manage complex, nested scientific workflows. These are a core ingredient of an OSP (see also Sect. 3.1) and are needed since the vast majority of materials properties are computed by a sequence of different calculations with nontrivial logic to control the choice of physical and numerical parameters, the dependencies, and the convergence loops and to perform error recovery. This logic is known to experienced computational scientists; encoding it in workflows not only simplifies the management of simulations when they must be repeated many times with different input choices but also enables automated computation of materials properties. The latter aspect is extremely beneficial first because these automatic workflows can be handed over to, e.g., experimentalists that can use them to quickly evaluate the properties they are interested to but also because it becomes possible to recompute the properties with different approximations or different codes for verification and validation purposes, to contribute to existing databases by adding new materials with their properties, and to validate the correctness of these new contributions automatically.

Finally, *Sharing* focuses on creating a social ecosystem to encourage data and workflows reuse and accelerate research discoveries. AiiDA can export all data stored, including the whole provenance graph and not only inputs and outputs. This information can be then imported in another AiiDA instance and used for new simulations or for further data analytics. AiiDA adopts common formats for typical data structures (like crystal structures, electronic bands, molecular-dynamics trajectories, … ) that are independent of the simulation codes. Thanks to the automatic tracking of the provenance, moreover, these can be converted in any other standard format at any time, as we discuss with an example in Sect. 4.3.
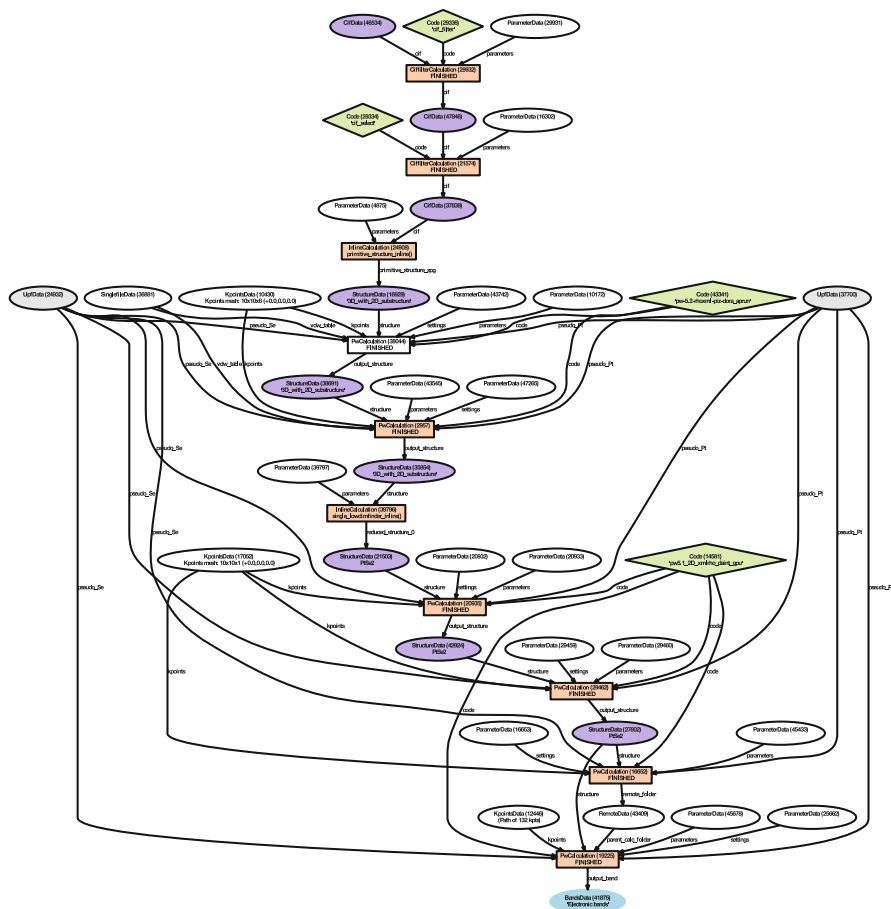
**Fig. 1** The provenance graph for the band structure of a 2D material (PtSe$_2$) from Mounet et al. (2018). AiiDA automatically records the provenance and can display it in graphical form with the `verdi graph generate` command. Here, calculations are shown as orange rectangles and data as ellipses. The codes used (binaries with their location, libraries) are also recorded in the provenance graph (green diamonds). We have highlighted with colors some specific data types: pseudopotentials in gray, the final band structure in blue, and crystal structures in purple. Data nodes in white are other input parameters for the calculations (e.g., k-point grids or input keywords). The graph makes it apparent that the final band structure for the 2D material was obtained starting from an initial 3D structure, in this case from the Crystallography Open Database (COD) (top purple node), first via a set of filtering steps, then via a Quantum ESPRESSO relaxation, followed by the geometrical exfoliation and a final relaxation and band structure calculation for the 2D structure

## 2.3    ADES in Open Science

We believe that one of the most effective approaches in the implementation of the ADES model, especially in the context of an OSP, is to couple automation and data storage. More specifically, AiiDA requires that all information needed to generate the calculation inputs is stored in the database even before the calculation is launched. The daemon is then responsible to create the actual input files, using only information already present in the database.

This coupling comes at a small cost for the researchers when new calculations are generated, as they must provide inputs using the AiiDA data structures and run them through the daemon (even if the additional barrier can be lowered by simplifying the user interface, as prescribed by the environment pillar). On the other hand, this approach ensures that common data structures are easily reusable even in different codes and that all calculations are reproducible and the provenance metadata is always correct. In fact, if the user had to add provenance information only after the execution of thousands of simulations, it would be a huge and tedious work, especially if calculations are not homogeneous. For this reason, in most cases researchers end up not adding complete provenance information to their data, unless strictly necessary. Moreover, this approach is error-prone as, e.g., the wrong inputs could be assigned to a given calculation.

The coupling of automation and data, simplified in a proper environment, is then strengthened (especially in view of our discussion on Open Science) by the sharing capabilities of data, provenance, and workflows. In the next sections, we will show with a number of examples how an ADES tool becomes an essential ingredient of an OSP and completes it when coupled with additional services.

## 3    Open Science Pillar 2: Making It Accessible

### 3.1    Not Only Data: Sharing of Workflows and Plugins

Sharing research data is one of the components of Open Science, as it allows other researchers to reproduce and reuse the results, perform new data analytics, and start new research using the data published. Hence, data sharing in an open and reusable format is essential. To this aim, recently, in the materials science community, a number of sharing portals have appeared, some of which are also discussed in this Handbook, like the Materials Project, AFLOWlib, OQMD, Nomad, and the CMR.

We emphasize, however, that sharing should not only be made easy and with appropriate open licenses, but also it is essential that the data is distributed together with sufficient metadata information to understand it and how it was generated. If simulations are run with an ADES tool like AiiDA that tracks data provenance, this aspect is simply addressed by always sharing data together with the corresponding provenance. Furthermore, data and metadata should be shared in a standard format.

Defining standard formats is a community effort and cannot be done by a single research group. Nevertheless, automatic provenance tracking makes it possible to export data after its generation in any other format, as discussed in Sect. 4.3, and hence any future standard can also be supported with very limited effort.

However, we believe that Open Science cannot be limited to open data sharing, as reproducibility is hindered if the tools used to generate the data are not available. These tools include the quantum simulation packages to compute materials properties as discussed already in Sect. 2.1 but also the pre- and post-processing tools to analyze the data and more generally the scientific workflows to obtain a set of materials properties from an initial minimal input, like a crystal structure. These are composed of a number of components: the infrastructure to run the workflows (like AiiDA), the plugins to interface AiiDA with the various codes adopted, and the logic that encodes the scientist's knowledge on the choice of numerical and physical parameters, on how to perform the sequence of calculations, and on how to deal with potential errors or convergence issues. Only if all these components are openly available, we can truly speak of reproducible Open Science. Indeed, we emphasize here that the availability of the provenance graph for a given dataset like the one shown in Fig. 1 is extremely useful but often allows only to reproduce that single result. In many cases, however, the graph is not enough to understand the sophisticated workflow logic used to select input data (like cutoffs or numerical parameters) or to filter relevant results (e.g., in a high-throughput study).

The challenge for Open Science, encoded in the second pillar discussed in Sect. 1.1, consists then in ensuring that all these components can be shared and reused with limited effort, that they can be interoperable and used together, and that contributions to them by third-parties with extensions or improvements are encouraged.

## 3.2 Encouraging Contributions

An OSP should be general and support a variety of simulation codes, data types, and workflows. Clearly, maintaining and supporting this ecosystem cannot be sustained by a single group, in particular for plugins to support specific simulation codes, because writing them requires in-depth knowledge of the code and of its typical usage patterns. Even more, this is true for workflows that are both coupled to the codes used and to the specific research field or topic.

As a consequence, in AiiDA we opted for a plugin interface. The main infrastructure, "AiiDA core," only contains the main logic that is independent of the codes, like dealing with external supercomputers, storing data and provenance in a database, or querying it. All the tools specific to codes, data, and workflows are implemented as plugins in independent repositories. A design based on plugins is essential, but an effective implementation can facilitate their installation and encourage contributions. For instance, in earlier versions of AiiDA, plugins had to be contributed to the code repository of AiiDA. Since version 0.9, the AiiDA plugin infrastructure has been improved: these can now be developed in independent repositories, and final users can easily choose which plugins to install. Once

installed, AiiDA automatically detects and uses them. The advantage is twofold: first, it encourages researchers to contribute their own code, as they do not lose control over it but maintain full authorship and can even decide a custom licensing scheme. Most importantly, if plugins live in different repositories, their development can occur independently without having to tie and synchronize their releases (e.g., if one code just needs a bugfix while another one is in the process of a big refactoring and is not ready for a new release).

During the design and improvement of the plugin interface, the AiiDA team soon recognized the need of a centralized repository to list existing plugins. In fact, in the past few years, it occurred that more than one researcher started to develop a plugin for the same code, driven by their needs. As a result, two very similar but essentially incompatible plugins were released. This results in work duplication with the additional risk that both plugins miss some important feature present in the other one. As a consequence, users that are faced with the choice of a plugin might get confused. Moreover, having multiple, slightly different formats also hinders sharing and reusability. To address this issue, the AiiDA plugin registry (AiiDA Team 2017a) has been created (see also Fig. 2), consisting in a centralized list of existing plugins, brief notes on how to use them, and links to their code repository and documentation. The plugin code is not copied or duplicated. Instead, the registry acts merely as an index to facilitate the discovery of existing plugins. Anybody can register a new plugin, and developers are encouraged to do so in the very early stages of development. Besides allowing them to reserve the plugin name (that needs to be unique among all plugins and should not be changed over time), this policy also reduces the risk that multiple researchers start independently to develop plugins for the same code. To describe the readiness of plugins, a "state" flag mentions if the plugin is stable and ready for production, under development, or only registered. Finally, a "plugin-template" repository is also provided (and is also available on the plugin registry) that can be copied and modified to start developing a new plugin very easily. To prove the effectiveness of this approach, we note that just 1 year after the creation of the registry, 22 different plugins are already available, including plugins for widespread simulation codes like Quantum ESPRESSO, VASP, CP2K, FLEUR, SIESTA, and YAMBO. Moreover, more than half of them provide detailed documentation websites, also in this case facilitated by the backbone documentation structure provided by the plugin template.

Finally, also in the spirit of facilitating usage and contributions and of increasing interoperability, AiiDA has been integrated and made compatible with other libraries and tools written in python that is becoming an extremely popular programming language in the materials science community. For instance, AiiDA can import and export structures between its internal format and the format of ASE (Larsen et al. 2017) and pymatgen (Ong et al. 2013), so that functionality present in these two libraries (crystal structure manipulation, comparison, processing) can be seamlessly integrated in any AiiDA workflow. Similarly, crystal symmetry can be analyzed using spglib (Togo and Tanaka 2018), and high-symmetry $k$-points in the Brillouin zone (together with suggested paths for band structures) can be automatically obtained via seekpath (Hinuma et al. 2017) and directly used as input to calculations.

**Fig. 2** Left, a screenshot of the homepage of the AiiDA plugin registry discussed in Sect. 3.2 and available at http://aiidateam.github.io/aiida-registry, that contains a list of all available plugins for AiiDA. Right, an example of a detail page for one of these plugins, containing general information on which functionality is provided by the plugin, how to install and use it, and where to find the code and the respective documentation

## 3.3 Documentation, Tutorials, and Open Learning Resources

The availability of open codes and tools for data generation and analysis is, as described in the previous section, an essential aspect of Open Science. If, however, these tools are not straightforward to use, their adoption is strongly hindered. We believe therefore that it is essential to invest resources in lowering the barrier for users. To achieve this goal, it is needed to have well-written and extensive documentation for the codes, describing how to install them and how to specify their inputs. In our experience, it is important to ensure that the documentation is easily searchable and indexed in web engines. Moreover, it is very effective to provide quick-start guides, example scripts or input files that can be copied and adapted by users, and a "cookbook" of useful recipes that are typically repeated by many users, like how to achieve a particular task or how to troubleshoot potential problems. In our experience, we found the use of a notebook-like format particularly effective (in particular with the use of jupyter, that we describe in more detail in Sect. 4.4 with some examples of its use in the Materials Cloud platform). On top of this, documentation specifically designed for developers is greatly useful for contributors to understand the code and be able to independently add new features and is therefore another tool to encourage contributions to the platform (as discussed in Sect. 3.2).

However, documentation alone (while essential) can easily become overwhelming for large codes and therefore might not lower enough the entrance barrier. Also tutorials and schools are extremely powerful tools for code dissemination. These events put researchers in direct contact with the developers of the tools, helping in creating a network and direct communication channels. The advantage is on both sides: users can get direct benefit by faster and more effective learning, and developers can profit from these events to know how their software is used, to collect useful feedback on common use-cases and to understand how to improve the code usability.

Finally, we also believe that it is important to provide access to educational resources like recordings of lectures, tutorials, and schools that cover both the basics of the science of the field as well as the code usage and applications. These resources should be disseminated in an open format, providing access to all students and young researchers. An example of this is shown in Sect. 4.4, where we discuss the learn section of Materials Cloud, a hub for educational videos and resources in the field of materials science.

## 3.4 Virtual Machines and the Quantum Mobile

Researchers often have a specific problem to address and look for a software that can solve it. A good documentation and effective tutorials can help convince them of the functionality of the code, and jupyter notebooks can facilitate its use, but the main initial barrier of installation and configuration remains. Unfortunately, the

potential issues that the users can experience are quite varied, are very dependent on their computer configuration, and are often related to code dependencies (need for C/fortran libraries, specific compilers or python packages; operating system incompatibilities; conflicts between multiple library versions; ... ). If the installation turns out to be too complex, the researchers might be discouraged (especially if they are not certain that the code can help them). The situation can be further complicated by the need of using more than on tool at the same time and making them interact.

To address this barrier, we suggest a few different approaches to reduce and almost remove the time required to start using a software (or a collection of softwares).

The simplest solution that does not require setting up any online server is the creation of virtual machines with useful software preinstalled. The distribution is very simple, consisting in sharing a single file (typically of a few GB in size) with the virtual machine image. Once this file is obtained, by just running the virtual machine, users find all codes preinstalled and preconfigured and can directly start to run simulations. A practical example of this is the Quantum Mobile (AiiDA Team 2017b) virtual machine. It contains a number of simulation codes preinstalled (including Quantum ESPRESSO, FLEUR, SIESTA, Yambo, CP2K, Wannier90) as well as AiiDA, preconfigured to run these codes within the virtual machine and coming already with a database preconfigured. This setup is ideal for education in computational materials science classrooms, where students can focus directly on the problem at hand and on understanding the results. This has been already proven in the 2018 "Molecular and Materials Modelling" course at ETH Zürich, in which the Quantum Mobile has been used as the platform to run the simulations, only after less than 4 months after its first release.

For larger simulations, running within the virtual machine is not ideal or even possible because of CPU or memory limitations, but researchers can still use it by configuring AiiDA to connect to the supercomputers they have access to. An alternative virtualization solution includes the use of Docker that can be thought as a tool to create lightweight virtual machines, where the Linux kernel and some resources are shared and not emulated. AiiDA now comes with a Docker image preconfigured, and this has pushed also many of the codes mentioned before to provide their own Docker images. The Docker setup provides a very similar level of containerization and code setup reproducibility of a virtual machine and is much more lightweight (for instance, in terms of disk usage when multiple similar machines are executed). Therefore, it is suited for tasks where many equivalent systems need to be automatically created. A typical use-case is in continuous testing platforms. On the other hand, however, it is worth noting that the use of Docker images is less indicated for educational purposes with respect to virtual machines. In fact, virtual machines are (still at the time of writing) much easier to start up without advanced knowledge of computer administration and of Docker and its related technologies. Additionally, a virtual machine provides seamless access to GUI applications (like text editors or file browsers), while with Docker one would need to access the instance through the command line, creating a potential additional barrier for students.

## 3.5    Supercomputer Centers

A second approach to facilitate users in the adoption of codes is to deploy them in supercomputer centers where the simulations are run. Already now most high-performance computing (HPC) centers have an application support team that compiles various versions of the codes used by their users and provides them as modules. The same could be done for tools to manage simulations like AiiDA. This might require, however, that these centers transition from a classical model of HPC providers to more general service providers for, e.g., virtual machines, database services, or long-term storage. A number of centers (like, in Europe, the Swiss CSCS, the Italian CINECA, or the German JSC) are already in the process of providing these new services and working together to federate access to them; we expect that this trend will be followed by even more centers in the next few years.

## 4    Open Science Pillar 3: Seamless Data Sharing and Preservation

### 4.1    Data Management Plans and FAIR Compliance

In the first phases of research, offline tools and codes are typically used to perform and organize research, as the data generated can be large and is typically confidential. However, when data must be shared, as described in the third pillar of an OSP described in Sect. 1.1, it becomes essential to employ suitably designed web portals. In addition to this, currently many funding agencies require to comply with data management plans (DMPs) for data dissemination and long-term preservation. Ideally, data should be compatible with the FAIR principles of sharing (Wilkinson et al. 2016) that require data to be findable, accessible, interoperable, and reusable. Findability can be achieved only if research data is associated with persistent handles like DOIs to make it citable. Free portals exists, like Zenodo (2018), that assign DOIs to datasets obtained from research projects and also guarantee long-term preservation. However, while files generated by the different codes can be uploaded on these services, this would still require that another researcher has a compatible software installed in order to open and analyze the results in the format uploaded by the original author. To remove this barrier, web portals need therefore to address also the other FAIR aspects to become effective OSPs.

In the next few sections, we show examples of how the goals of being DMP and FAIR-compliant can be conveniently achieved using a combination of suitable web portals together with an ADES tool like AiiDA to manage simulations and track provenance. We discuss in particular the integration of AiiDA with the Theoretical Crystallography Open Database (TCOD) in Sect. 4.3 that proves how the automatic tracking of data provenance makes it possible to tag results with standard metadata automatically and after the simulations have run; and the

Materials Cloud portal (Sect. 4.4) that provides tools encompassing all steps of computational research, from learning to generating data and curating the results, to finally publishing and sharing results, while being fully compliant with the FAIR principles.

## 4.2 Interoperability Between Different Databases: The OPTiMaDe API

Even if automated provenance tracking allows to store a posteriori the metadata information in any format, as we discuss in the next section, it is still extremely valuable that standards are defined. This, in fact, would allow easy interoperability between different existing databases. A community effort in the direction of database and web-portal interoperability is happening in the OPTiMaDe consortium that is working toward the definition of a common API specification to be implemented by the different partners. Many of the large databases in the community are already part of this consortium (including Materials Cloud that has the _mcloud_ prefix assigned to it). In the current version (0.9.5), OPTiMaDe already defines a REST API format that makes it possible to query, with the same format, for the existence of crystal structures in different databases, with common filters like number of atoms, presence or absence of a chemical element, as well as supporting database-specific fields.

## 4.3 Automatic a Posteriori Metadata Tagging: AiiDA and TCOD Integration

Ontologies and other standards for metadata tagging and sharing of research results in materials science are currently being discussed in the community, like, e.g., the TCOD dictionaries (Gražulis et al. 2014) and the Nomad metadata (Nomad Meta Info 2018), even if there is not an established standard yet. For this reason, it is difficult to enforce a given ontology. The lack of a single standard format in the community, however, is not a major issue if simulations are run with automatic provenance-tracking tools. Indeed, tracking of provenance (if complete and automatic) allows users, a posteriori, i.e., after all simulations have run, to convert the provenance information in any other format by just implementing a converter. An example of this is shown in Merkys et al. (2017), where methods and codes have been presented to convert the provenance as tracked by AiiDA to the format defined by the TCOD ontology. In this work the authors show in particular how Quantum ESPRESSO simulations managed via AiiDA are automatically tagged with metadata using to the TCOD ontology, with no user input required. Moreover, the implementation is modular and additional plugins can be developed to support other simulation tools or other metadata formats.

## 4.4   Materials Cloud

In Sect. 2 we have described how AiiDA addresses the challenges of data generation in an OSP, in Sect. 3 we have outlined how to make Open Science accessible, and in Sect. 4.1 we have emphasized how web portals tailored to comply with the FAIR principles are an essential ingredient to achieve seamless sharing.

For this reason, Materials Cloud has been designed, implemented, and deployed. It is a web portal for materials science that is coupled to AiiDA and, in combination with it, addresses all three OSP pillars introduced in Sect. 1.1. Materials Cloud is composed of different sections that aim at assisting researchers during the full lifecycle of a scientific project.

The first section, *Learn*, contains educational material like videos of schools, tutorials, and lectures in the field of materials science, together with the corresponding material to assist learning (like slides and exercises). The learn section uses the SlideShot technology, a software platform that makes it possible to show the video feed of the speaker together with the slides in high resolution, allowing viewers to quickly seek through the video via the slides thumbnails. This is coupled with a standard hardware setup to perform recordings and then import them into SlideShot.

The *Work* section then focuses on the task of data generation. This section addresses in particular the second pillar of Open Science to make the simulation tools available in the form of automated solutions, removing access barriers and making them available not only to experienced computational researchers but also to experimentalists or to students. Besides providing links to download various versions of the Quantum Mobile virtual machine (see Sect. 3.4), it provides access to computational tools with two different approaches, suitable for different use-cases that we briefly discuss here. The first approach of the work section of Materials Cloud is to provide a set of *online tools* to perform fast analysis of data, directly from the web browser, similar in the spirit to, e.g., the many tools of the Bilbao Crystallographic Server (Aroyo et al. 2011). This is particularly suited for computations that can run in real time (i.e., with a running time of up to a few seconds) and that can benefit from a graphical web interface. An example is seekpath (Hinuma et al. 2017), a tool that, given an input structure, computes a standardized primitive unit cell according to the standard definitions in the crystallography literature, like in the International Tables of Crystallography (Shmueli 2010) and in Parthé and Gelato (1984). The tool then returns also the labels of high-symmetry points in the Brillouin zone together with a suggested path for the computation of band structures. The only required input, in this case, is a crystal structure (accepted in a number of common formats) and, optionally, a few numerical parameters; interactivity happens mainly via to interactive 3D visualizers for the crystal structure and the Brillouin zone.

The second and more flexible approach for data generation in the Materials Cloud is based on the *jupyter* interface. Jupyter is a notebook-like web front end that allows to run any python code (and also supports many other programming languages) subdivided in cells with inputs and outputs. This is a very flexible interface, and in Materials Cloud it is further powered by the AppMode plugin that by default hides
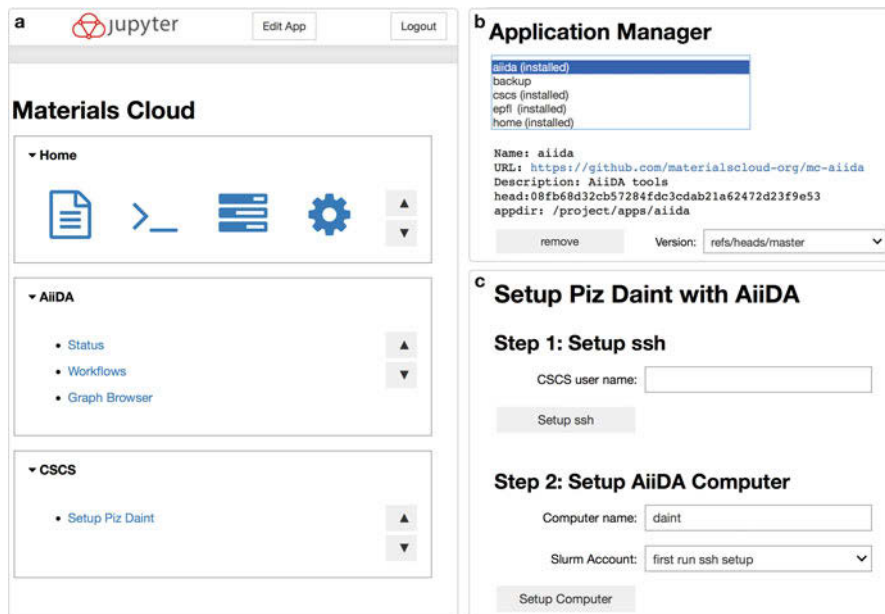
**Fig. 3** Three screenshots of the jupyter interface of the Materials Cloud. Pages are customized with the AppMode plugin, to make the appearance very similar to a standard web page. (**a**) Home page, with the top panel with main tools to manage Materials Cloud applications, to show a terminal or to show the control panel. On the top, a "Edit app" button allows to go back to the standard jupyter notebook interface. (**b**) Application manager that allows to select existing apps from the Materials Cloud app registry and make them appear in the home page. (**c**) One of the apps to setup a new computer (in this case, the Piz Daint supercomputer at the Swiss Supercomputer Centre) with just a few clicks

the cells and just shows the outputs (including widgets like buttons, text boxes, and drop-down lists) in a format that resembles a standard interactive web page. The jupyter interface of the Materials Cloud is shown Fig. 3, providing a home page where contributed apps can be added directly via the web interface (panels a and b). Apps can be provided by anyone and just need to be registered in the Materials Cloud App registry. Users are provided with a working space already preconfigured with AiiDA and codes, removing any setup time. Additional configuration (e.g., setting up AiiDA to interact with custom computational clusters) is significantly simplified by apps with simple GUIs (see Fig. 3c). The only disadvantage of this section is that it requires a user login, because users can access a full computer with unrestricted access to code execution and internet access, but on the other hand, it is the ideal platform to create custom fully automated workflow solutions for the computation of materials properties.

Once the data has been generated (locally with AiiDA on in the jupyter section), it can be displayed in the two sections *Discover* (for curated data) and *Explore* (for the "raw" data as generated via AiiDA). The first section can be used by a

researcher to provide an accessible interface to understand a project and present its results with dynamic data filtering, as well as interactive views of the figures shown in the corresponding papers. Data can be linked to the corresponding AiiDA nodes in the explore section. In the latter, data is visualized together with its provenance (browsable via an interactive graph to inspect the calculation that generated it or to find out by which calculations it was used as input). Calculations are always displayed together with their full set of inputs and outputs that can also be directly viewed or downloaded. Materials Cloud has been designed as an extension of AiiDA so that, if the simulations are run with AiiDA, the explore section is filled in automatically by just importing an AiiDA export file. To make an analogy: AiiDA is a tool used to manage simulations and provenance locally on a computer and can be compared to Git, used locally for the organization and tracking of the history of files and source codes. Then, the Materials Cloud explore section plays the role of GitHub, GitLab, or similar web services, providing web browsing of the content of repositories and acting as a central server for sharing.

Finally, Materials Cloud has a fifth section, *Archive*, for the long-term storage and dissemination of research results. A DOI is assigned automatically to each submitted entry, and standardized metadata are exposed in the XML Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) format (XML OAI-PMH 2018). Each entry can contain files with research results and can be linked to the corresponding discover and explore sections. An example screenshot of an entry of the Materials Cloud Archive is shown in Fig. 4.

By coupling the different sections, Materials Cloud realizes a FAIR-compliant platform as discussed in Sect. 4.1. In fact, the DOIs assigned to published research data in the archive section achieve findability. At the same time, if the data has been generated with AiiDA, each entry can be linked to a curated discover section and to an explore section, making it possible to visualize inputs and outputs of calculations and more generally to explore the data provenance. This real-time web interface makes data accessible (second FAIR pillar) and does not create any barrier due to authentication or to software installation. Interoperability (third FAIR pillar) is achieved since data that is common to different codes (e.g., crystal structures, electronic and phonon bands or $k-$point sets) are stored in a code-independent format and the outputs of a calculation can be used as the inputs for a different code. Finally, reusability (fourth FAIR pillar) is guaranteed by the possibility of getting all inputs and outputs of calculations and reproduce them, coupled with the large selection of licenses that the users can choose (with encouragement for those allowing for reuse, like the Creative Commons ones).

## 5    Examples of Open-Science Research Using AiiDA and the Materials Cloud

To better explain the ideas and tools described in the previous sections, we briefly discuss here two scientific projects, managed with AiiDA and hosted on the Materials Cloud, that demonstrate the concepts of an OSP discussed here.

A Standard Solid State Pseudopotentials (SSSP) library optimized for accuracy and efficiency (Version 1.0, data download)

Authors: Gianluca Prandini[1*], Antimo Marrazzo[1], Ivano E. Castelli[1,2], Nicolas Mounet[1], Nicola Marzari[1*]

1 Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

2 Department of Energy Conversion and Storage, Technical University of Denmark, Fysikvej 309, 2800 Kgs Lyngby, Denmark

* Corresponding authors emails: gianluca.prandini@epfl.ch, nicola.marzari@epfl.ch

(version: v1, submitted on: 26 January 2018)

How to cite this entry

Gianluca Prandini, Antimo Marrazzo, Ivano E. Castelli, Nicolas Mounet, Nicola Marzari, *A Standard Solid State Pseudopotentials (SSSP) library optimized for accuracy and efficiency (Version 1.0, data download)*, Materials Cloud Archive (2018), doi: 10.24435/materialscloud:2018.0001/v1.

Description

Despite the enormous success and popularity of density functional theory, systematic verification and validation studies are still very limited both in number and scope. Here, we propose a universal standard protocol to verify publicly available pseudopotential libraries, based on several independent criteria including verification against all-electron equations of state and plane-wave convergence tests for phonon frequencies, band structure, cohesive energy and pressure. Adopting these criteria we obtain two optimal pseudopotential sets, namely the Standard Solid State Pseudopotential (SSSP) efficiency and accuracy libraries, tailored for high-throughput materials screening and high-precision materials modelling. As of today, the SSSP accuracy library is the most accurate open-source pseudopotential library available. This archive entry contains the database of calculations (phonons, cohesive energy, equation of state, band structure, pressure, etc.) together with the provenance of all data and calculations as stored by AiiDA.

Materials Cloud sections using this data

⚛ SSSP Discover page
⁂ SSSP Explore page

Files

| File name | Size | Description |
|---|---|---|
| SSSP_efficiency_pseudos.tar.gz<br>MD5 | 34.4 MiB | Standard Solid State Pseudopotentials Efficiency for high-throughput materials screening v1.0. |
| SSSP_accuracy_pseudos.tar.gz<br>MD5 | 35.0 MiB | Standard Solid State Pseudopotentials Accuracy for high-throughput materials screening v1.0. |
| SSSP_efficiency_pseudos.aiida<br>MD5 | 34.4 MiB | Standard Solid State Pseudopotentials Efficiency for high-throughput materials screening v1.0 (AiiDA export file). |

**Fig. 4** An entry of the Materials Cloud Archive, in this case version 1 of the SSSP pseudopotential library (Prandini et al. 2018a). Each entry, beside typical metadata like authors, affiliations, and a description, can also contain a number of files that are preserved in the long term. Moreover, entries can have direct link to respective discover or explore sections, where data is directly visualizable in the web browser. Materials Cloud assigns each entry a DOI (the prefix of Materials Cloud is 10.24435) and requires authors to choose a license

## 5.1   The SSSP Pseudopotential Library

The Standard Solid State Pseudopotentials (SSSP) (Prandini et al. 2018b) library provides a curated selection of pseudopotentials for plane-wave density-functional theory (DFT) codes (in UPF format). SSSP is composed of two sublibraries, optimized, respectively, for *precision* and *efficiency* by means of a number of convergence studies on elemental solids, for various relevant physical properties including zone-boundary phonons, cohesive energy, pressure, and band structure.
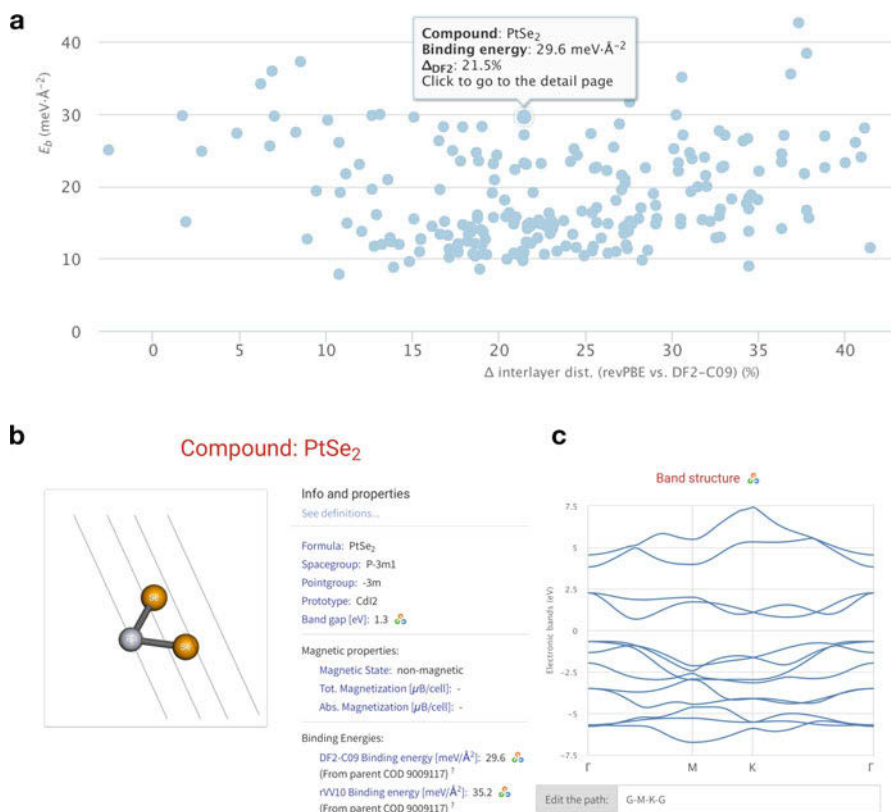
**Fig. 5** Screenshots of the discover section for the exfoliable 2D materials section. (**a**) One of the three selection pages, allowing to select a 2D material based on its binding energy and the change in interlayer distance in the bulk when computed with a DFT energy functional including or not including van der Waals force contributions. Other selection views include a periodic table and a detailed table. (**b**) Top part of the detail view for one of the materials (in this case, $PtSe_2$). The view includes an interactive structure 3D visualizer, as well as the main properties computed for this material. (**c**) Another portion of the detail view showing the electronic band structure of $PtSe_2$. The plot is interactive and zoomable, and the default path (here $\Gamma-M-K-\Gamma$) can be changed by the user. Both here and in panel **b**, the small AiiDA icons are links that bring the user to the corresponding explore section, to browse the provenance of the corresponding data (the provenance of the band structure is the one shown in Fig. 1)

The pseudopotential library is available on the Materials Cloud Archive (Prandini et al. 2018a); see also Fig. 4.

In the respective discover section, a periodic table uses a color legend to indicate the optimal pseudopotential for each element and indicates the suggested cutoff values. Clicking on an element shows a detailed view containing all convergence studies and plots (equations of states, band structures, bands chessboards) for all the pseudopotentials compared in the study. Datapoints in the interactive plots are clickable and bring the user to the explore section, with browsable provenance for

all data. Simulations were run using Quantum ESPRESSO and managed by AiiDA that also tracked the provenance.

## 5.2 The Exfoliable 2D Materials Database

In Mounet et al. (2018), the authors performed a high-throughput search of novel two-dimensional (2D) materials by screening the ICSD and the COD databases. Starting from over 100,000 unique bulk experimentally known materials, they could first detect with a geometrical algorithm the ∼6000 that are composed by layers and then refine this list with DFT binding-energy calculations (using Quantum ESPRESSO and managed by AiiDA) to only include those held together by weak forces. With this approach they could identify a set of about 1800 potentially or easily exfoliable 2D materials. Furthermore, a subset of 254 materials (easily exfoliable with up to 6 atoms per unit cell) has been extensively studied to compute relevant electronic, vibrational, and magnetic properties. These results and calculations are available on the Materials Cloud Archive (Mounet et al. 2017). Also in this case, the archive section is coupled to a curated discover page (Fig. 5). In the filtering page, a 2D structure can be selected via a table, by selecting elements in a periodic table or by picking a material from a binding-energy plot. After having selected the material, a detailed view is presented, showing data and results that include the binding energy of the 2D layer, the interlayer distance computed with different van der Waals functionals, the magnetic ground state (also including antiferromagnetism in larger supercells), the electronic band structure, and the phonons. These were computed with the correct 2D physics that properly considers electrostatic screening in low-dimensional systems, reproducing correctly the behavior of longitudinal and transverse optical phonons near $\Gamma$ (Sohier et al. 2017). Final results (plot and numerical values) are accompanied by small AiiDA icons, as shown in Fig. 5b, c. These are links to the corresponding explore page, where users can check how the results have been computed by browsing their AiiDA provenance.

## 6 Conclusions

In this chapter we have introduced our vision for an Open-Science platform. Such a platform should rest on three main pillars, namely: open data generation tools, open integration platform, and seamless data sharing. We have discussed the challenges set up by these requirements, and we have shown how a combination of the AiiDA code and the Materials Cloud platform can achieve the goals defined by these three pillars. These aspects have furthermore been demonstrated with two examples of projects that used the two tools, namely, the SSSP pseudopotential library and the database of exfoliable 2D materials. Indeed, if simulations are run with AiiDA, the generated data is reproducible and its provenance (i.e., how it was generated) is automatically tracked and stored. AiiDA also helps creating and steering automatic workflows for the calculation of materials properties. Combined with the Materials

Cloud work section (and in particular with the jupyter subsection), advanced workflows can be exposed with an intuitive and easy interface, minimizing the barrier to access and use them. The other sections of Materials Cloud, furthermore, enable sharing of computed data in a FAIR-compliant format. DOIs are assigned to data entries (that are preserved in the long term) to make them findable, as required by data management plans. These entries, moreover, are made accessible by a tight coupling with interactive views to present curated data. The latter is linked to browsable graphs to access and explore the data provenance automatically tracked by AiiDA as well as the raw input and output files. AiiDA and Materials Cloud, therefore, implement our Open-Science vision with the aim of making science accessible to everybody and of encouraging reuse of results to promote and support scientific discovery.

# References

AiiDA Team (2017a) Quantum mobile. https://www.materialscloud.org/work/quantum-mobile
AiiDA Team (2017b) The AiiDA plugin registry. https://aiidateam.github.io/aiida-registry/
Aroyo MI, Perez-Mato JM, Orobengoa D, Tasci E, de la Flor G, Kirov A (2011) Crystallography online: bilbao crystallographic server. Bulg Chem Commun 43(2):183–197
Belsky A, Hellenbrandt M, Karen VL, Luksch P (2002) New developments in the inorganic crystal structure database (ICSD): accessibility in support of materials research and design. Acta Cryst B 58(3 Part 1):364–369. https://doi.org/10.1107/S0108768102006948
Blügel S, Bihlmayer G (2006) Full-potential linearized augmented planewave method. In: Computational nanoscience: do it yourself!. NIC series, vol 31. John von Neumann Institut for Computing, Julich, p 85
Clark SJ, Segall MD, Pickard CJ, Hasnip PJ, Probert MIJ, Refson K, Payne MC (2005) First principles methods using CASTEP. Z Kristallogr 220:567–570. https://doi.org/10.1524/zkri.220.5.567.65075
Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, Nelson LJ, Hart GL, Sanvito S, Buongiorno-Nardelli M, Mingo N, Levy O (2012) AFLOWlib.org: a distributed materials properties repository from high-throughput ab initio calculations. Comput Mater Sci 58:227–235. http://www.sciencedirect.com/science/article/pii/S0927025612000687
Dragoni D, Daff TD, Csányi G, Marzari N (2018) Achieving dft accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron. Phys Rev Mater 2:013808. https://doi.org/10.1103/PhysRevMaterials.2.013808
Giannozzi P, Andreussi O, Brumme T, Bunau O, Nardelli MB, Calandra M, Car R, Cavazzoni C, Ceresoli D, Cococcioni M, Colonna N, Carnimeo I, Corso AD, de Gironcoli S, Delugas P Jr,

RAD, Ferretti A, Floris A, Fratesi G, Fugallo G, Gebauer R, Gerstmann U, Giustino F, Gorni T, Jia J, Kawamura M, Ko HY, Kokalj A, Küçükbenli E, Lazzeri M, Marsili M, Marzari N, Mauri F, Nguyen NL, Nguyen HV, de-la Roza AO, Paulatto L, Poncé S, Rocca D, Sabatini R, Santra B, Schlipf M, Seitsonen AP, Smogunov A, Timrov I, Thonhauser T, Umari P, Vast N, Wu X, Baroni S (2017) Advanced capabilities for materials modelling with quantum espresso. J Phys Condens Matter 29(46):465901. http://stacks.iop.org/0953-8984/29/i=46/a=465901

Gonze X, Jollet F, Araujo FA, Adams D, Amadon B, Applencourt T, Audouze C, Beuken JM, Bieder J, Bokhanchuk A, Bousquet E, Bruneval F, Caliste D, Côté M, Dahm F, Pieve FD, Delaveau M, Gennaro MD, Dorado B, Espejo C, Geneste G, Genovese L, Gerossier A, Giantomassi M, Gillet Y, Hamann D, He L, Jomard G, Janssen JL, Roux SL, Levitt A, Lherbier A, Liu F, Lukačevič I, Martin A, Martins C, Oliveira M, Poncé S, Pouillon Y, Rangel T, Rignanese GM, Romero A, Rousseau B, Rubel O, Shukri A, Stankovski M, Torrent M, Setten MV, Troeye BV, Verstraete M, Waroquiers D, Wiktor J, Xu B, Zhou A, Zwanziger J (2016) Recent developments in the abinit software package. Comput Phys Commun 205:106–131. https://doi.org/10.1016/j.cpc.2016.04.003, http://www.sciencedirect.com/science/article/pii/S0010465516300923

Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya NR, Moeck P, Downs RT, Le Bail A (2012) Crystallography open database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Res 40(D1):D420–D427. https://doi.org/10.1093/nar/gkr900

Gražulis S, Merkys A, Vaitkus A, Le Bail A, Chateigner D, Vilčiauskas L, Cottenier S, Björkman T, Murray-Rust P (2014) Launching the theoretical crystallography open database. Acta Cryst A 70(a1):C1736. https://doi.org/10.1107/S2053273314082631, http://www.crystallography.net/tcod

Hinuma Y, Pizzi G, Kumagai Y, Oba F, Tanaka I (2017) Band structure diagram paths based on crystallography. Comput Mater Sci 128:140–184

Hutter J, Iannuzzi M, Schiffmann F, VandeVondele J (2014) cp2k: atomistic simulations of condensed matter systems. Wiley Interdisc Rev: Comput Mol Sci 4(1):15–25. https://doi.org/10.1002/wcms.1159, https://doi.org/10.1002/wcms.1159

Jain A, Hautier G, Moore CJ, Ping Ong S, Fischer CC, Mueller T, Persson KA, Ceder G (2011) A high-throughput infrastructure for density functional theory calculations. Comput Matter Sci 50(8):2295–2310. http://www.sciencedirect.com/science/article/pii/S0927025611001133

Kresse G, Furthmüller J (1996) Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput Matter Sci 6(1):15–50. https://doi.org/10.1016/0927-0256(96)00008-0, http://www.sciencedirect.com/science/article/pii/0927025696000080

Landis D, Hummelshøj J, Nestorov S, Greeley J, Dułak M, Bligaard T, Nørskov J, Jacobsen K (2012) The computational materials repository. Comput Sci Eng 14(6):51–57

Larsen AH, Mortensen JJ, Blomqvist J, Castelli IE, Christensen R, Dułak M, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Jensen PB, Kermode J, Kitchin JR, Kolsbjerg EL, Kubal J, Kaasbjerg K, Lysgaard S, Maronsson JB, Maxson T, Olsen T, Pastewka L, Peterson A, Rostgaard C, Schiøtz J, Schütt O, Strange M, Thygesen KS, Vegge T, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW (2017) The atomic simulation environment—a python library for working with atoms. J Phys Condens Matter 29(27):273002

Marini A, Hogan C, Gröning M, Varsano D (2009) yambo: an ab initio tool for excited state calculations. Comput Phys Commun 180(8):1392–1403. https://doi.org/10.1016/j.cpc.2009.02.003, http://www.sciencedirect.com/science/article/pii/S0010465509000472

Materials cloud (2017) https://www.materialscloud.org

Merkys A, Mounet N, Cepellotti A, Marzari N, Gražulis S, Pizzi G (2017) A posteriori metadata from automated provenance tracking: integration of AiiDA and TCOD. J Cheminform 9(1):56. https://doi.org/10.1186/s13321-017-0242-y

Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, Sohier T, Castelli IE, Cepellotti A, Pizzi G, Marzari N (2017) Two-dimensional materials from high-throughput

computational exfoliation of experimentally known compounds (Data download). Mater Cloud Archive. https://doi.org/10.24435/materialscloud:2017.0008/v1

Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, Sohier T, Castelli IE, Cepellotti A, Pizzi G, Marzari N (2018) Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. Nat Nanotech 13:246–252. https://doi.org/10.1038/s41565-017-0035-5

Nomad Meta Info (2018) https://metainfo.nomad-coe.eu/nomadmetainfo_public/archive.html

Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier VL, Persson KA, Ceder G (2013) Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. Comput Matter Sci 68(0):314–319. https://doi.org/10.1016/j.commatsci.2012.10.028, http://www.sciencedirect.com/science/article/pii/S0927025612006295

Parthé E, Gelato LM (1984) The standardization of inorganic crystal-structure data. Acta Cryst A 40(3):169–183. https://doi.org/10.1107/S0108767384000416

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) Aiida: automated interactive infrastructure and database for computational science. Comput Matter Sci 111:218–230. https://doi.org/10.1016/j.commatsci.2015.09.013

Prandini G, Marrazzo A, Castelli IE, Mounet N, Marzari N (2018a) Precision and efficiency in solid-state pseudopotential calculations, npj Comput Mater 4:72

Prandini G, Marrazzo A, Castelli IE, Mounet N, Marzari N (2018b) A standard solid state pseudopotentials (SSSP) library optimized for accuracy and efficiency (Version 1.0, data download). Mater Cloud Archive. https://doi.org/10.24435/materialscloud:2018.0001/v1

Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA (2014) Quantum chemistry structures and properties of 134 kilo molecules. Sci Data 1:140022

Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C (2013) Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). JOM 65(11):1501–1509. https://doi.org/10.1007/s11837-013-0755-4

Shmueli U (ed) (2010) International tables for crystallography volume B: reciprocal space, 2nd edn. International Union of Crystallography. https://doi.org/10.1107/97809553602060000108

Sohier T, Gibertini M, Calandra M, Mauri F, Marzari N (2017) Breakdown of optical phonons' splitting in two-dimensional materials. Nano Lett 17(6):3758–3763. https://doi.org/10.1021/acs.nanolett.7b01090

Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, Sánchez-Portal D (2002) The SIESTA method for ab initio order-N materials simulation. J Phys Condens Matter 14(11):2745

The Nomad Repository (2018) https://repository.nomad-coe.eu

Togo A, Tanaka I (2018) Spglib: a software library for crystal symmetry search. arXiv e-prints: 1808.01590. http://adsabs.harvard.edu/abs/2018arXiv180801590T, https://atztogo.github.io/spglib/

Van Noorden R, Maher B, Nuzzo R (2014) The top 100 papers. Nature 514:550–553. https://doi.org/10.1038/514550a

Villars P, Berndt M, Brandenburg K, Cenzual K, Daams J, Hulliger F, Massalski T, Okamoto H, Osaki K, Prince A, Putz H, Iwata S (2004) The Pauling file, binaries edition. J Alloys Comput 367(1–2):293–297. https://doi.org/10.1016/j.jallcom.2003.08.058

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The fair guiding principles for scientific data management and stewardship. Sci Data 3:160018 EP –. https://doi.org/10.1038/sdata.2016.18

XML OAI-PMH (2018) http://www.openarchives.org/pmh/

Zenodo (2018) https://zenodo.org

# The PAULING FILE Project and Materials Platform for Data Science: From Big Data Toward Materials Genome

# 84

Evgeny Blokhin and Pierre Villars

## Contents

**Abstract**

One of the oldest initiatives in materials informatics, the PAULING FILE project, is described. It includes the comprehensive database for inorganic crystalline compounds, their atomic structures, intrinsic physical properties and phase

E. Blokhin (✉)
Materials Platform for Data Science, Tallinn, Estonia

Tilde Materials Informatics, Berlin, Germany
e-mail: eb@tilde.pro

P. Villars
Material Phases Data System, Vitznau, Switzerland
e-mail: villars.mpds@bluewin.ch

diagrams. On top of that, the powerful online retrieval software is introduced, called MPDS, the Materials Platform for Data Science. The practical recipes of storage, exchange, and analysis of the large amounts of materials data are given. The focus is made on the modern information technologies and software engineering. As a result, from the large heterogeneous data, holistic conclusions about the entire set of known materials are drawn. They can be regarded as a guideline for the systematic large-scale predictions.

## 1  Introduction

The PAULING FILE project is the materials database with nearly 25 years history, grouping crystallographic data, phase diagrams, and physical properties of inorganic crystalline substances under the same frame. Its focus is put on the experimental observations. Each individual crystal structure, phase diagram, or physical property originates from a particular peer-reviewed publication. The world's scientific literature in materials science, engineering, physics, and inorganic chemistry is covered from 1891 to the present date. The Materials Platform for Data Science is an online edition of the PAULING FILE project, created in 2016. It presents all the PAULING FILE data in two online interfaces: graphical user interface and application programming interface. The former is intended for the materials scientists; the latter is intended for the software engineers and data scientists. An intersection of the research interests of these groups falls into the scope of the novel discipline of materials informatics.

## 2  Materials Big Data and the PAULING FILE Project

### 2.1  Modern Challenges Ahead of the Materials Community

Counterintuitively, empirical traditions are widespread in the fundamental science and particularly in materials science. The pioneer in chemoinformatics, Peter Murray-Rust (Cambridge, UK), called the materials science as one of the most conservative precise disciplines, the least transparent and open to collaboration and crowd-sourcing, compared to the other sciences, as biochemistry, astronomy, mathematics, and computer and environmental sciences (Murray-Rust 2013) – not counting paywalls. Interesting to note, the field of materials informatics is less established and younger, compared to chemo- and bioinformatics (Gasteiger and Engel 2003, here *informatics* assumes an *information exchange*). Among the novel distributed computing projects, where anyone may volunteer computing time from their personal computers to a specific cause, only a very little part is concerned with materials science. The following not exhaustive explanation can be given. Materials science and engineering are maximally tightly bound to the industries, e.g., aerospace, automotive, electronics, military industry, etc. Considerable part of materials research and development is privately funded. A know-how in materials

science costs on average higher, than in other precise disciplines. These costs can be compared to high energy physics or astronomy. However an influence of such know-hows on the technological progress is also higher, and the lag between investment and innovation is relatively moderate (about 15 years or less, Obama 2011). Further, due to its widespread and extreme complexity, materials science is also very segmented ("feudal"). Consequently, the level of competition is very high, both in academia and industry (government grants, staff positions, etc.). In terms of complexity, even mastering the basic formalism of the modern solid-state physics presents a nontrivial task. The multidisciplinary specialists are generally rare in materials science. These reasons may give an idea, why the materials science community is inherently very conservative.

Moreover, until the recent years the culture of sharing the basic research outcome (e.g., raw measurement or simulation logs, in contrast to the articles) was totally absent and sometimes even tabooed in materials science. Concerning the publications, still there is no way to obtain scientific information in machine-readable format from journals for further reuse and repurposing. However with the technical progress the modern academics are sinking in the "ocean of data." Today there is a common complaint that researchers are publishing too much and too fast. To estimate this growth, it is wrong to cite the growth of academic databases, as no database captures everything. The bibliometric analysts from the Max Planck Society and Swiss Federal Institute of Technology (Bornmann and Mutz 2015) estimated that the global scientific output roughly doubles every 9 years. Moreover, with the increase in popularity of the modern data-intensive approaches, commonly denoted as "materials design" or "materials genome," there is an evidence that the amount of raw big data in materials science generated by experiments or simulations will continue to grow exponentially. Speaking about the *big data*, the computer science community has appropriately defined this phenomenon as the data amounts governed by four metrics: volume, velocity, variety, and veracity. In materials science *volume* refers to the big sizes, exceeding personal computer facilities, and *velocity* refers to harnessing real-time data acquisition (e.g., from dynamics experiments). *Variety* is concerned with the fact that the data takes all forms in materials science, ranging from discrete numerical values to qualitative descriptions of materials behavior and imaging data. *Veracity* acknowledges the practical reality about uncertainties and a lot of "missing" data (Rajan 2015). Nowadays is the epoch of big data in materials science. Yet in this epoch of big data, materials science is still not doing enough to encourage and enable the sharing, analysis, and interpretation of the vast swatches of data that researchers are collecting. The traditional means of exchange of scientific information in materials science community are deeply imperfect. In such conditions a natural diversity appears, with the principles of the natural selection guiding the shape of the cutting-edge research. And in the last several years with the high penetration of the Internet and modern information technologies, the situation has started to change slowly. Here the ability to digest information, drawing the correct conclusions, is crucial. This is where the data science tools (e.g., machine learning) need to be linked to the foundations of materials science: theory, modeling, and experiments. The aim is to

make the *laboratories* for generating the new information and not just *repositories* for retrieving the known or expected information (Ghiringhelli et al. 2017).

## 2.2 PAULING FILE Background

There are a number of initiatives in the world trying to overcome the above-mentioned challenges in materials science (Pizzi et al. 2016). The main idea is either to systematically collect the materials data or generate and process them in a high-throughput manner. One of the notable initiatives is the PAULING FILE project, launched in 1995. Its main focus is put on the critical evaluation of the published experimental materials data. Historically, it was a joint venture of the Japan Science and Technology Corporation, Material Phases Data System company in Switzerland, and the University of Tokyo, RACE. Now it is managed solely by the Swiss company. Three steps were planned from the beginning. The first goal was to create and maintain a comprehensive database for inorganic crystalline substances, covering crystallographic data, diffraction patterns, intrinsic physical properties, and phase diagrams. The data should be checked with the extreme care. The term "inorganic substances" was defined as compounds containing no C-H bonds. In parallel to the database creation, the second goal was to develop an appropriate retrieval software to make the data accessible in a single-user interface. In longer term, as the third goal, the new tools for materials design should be created, to search the database for correlations automatically. This is known as intelligent design of the new inorganic materials with predefined intrinsic physical properties. The pilot version PAULING FILE Binaries Edition was released as a desktop software in 2002. Now the selected parts of the PAULING FILE data are included in several printed, offline and online products. Today the PAULING FILE project is quite well-known. There are already thousands of publications referring it. Its foundations, database design, and data-centric observations are published (e.g., Villars 2004; Villars et al. 2008; Xu et al. 2011; Kong et al. 2012; Villars and Iwata 2013 etc.). The recent implementation of the PAULING FILE retrieval software and the materials design tools is an online product called Materials Platform for Data Science.

Now a minimum of required definitions must be given. A *database* is a collection of interrelated stored data that serves the needs of the multiple users. In the so-called relational paradigm, these stored data are organized in the *tables*. The motivations for using databases rather than files include greater availability to a diverse set of users, integration of data for easier access to and updating of complex transactions, and less redundancy of data. A *database management system* (DBMS) is a generalized software for manipulating the databases. A DBMS supports a logical view (*database schema*), physical view (access methods, data clustering), data definition and manipulation language, and utilities, such as transaction management and concurrency control, data integrity, crash recovery, and security (Teorey et al. 2005). For example, Oracle is used as an internal DBMS for PAULING FILE. The PAULING FILE database has the following structure. The standard unit of data is called an *entry*. All the entries are subdivided into three kinds: crystalline structures,

physical properties, and phase diagrams. They are called S-, P-, and C-entries, correspondingly. Entries have persistent identifiers, analogous to digital object identifiers (DOIs), e.g., S377634, P600028, and C100027. Another dimension of the PAULING FILE data is the *distinct phases*. The three kinds of entries are interlinked via the distinct materials phases they belong. A tremendous work has been done by PAULING FILE editors in the past 25 years to manually distinguish more than 150,000 inorganic materials phases, appearing in the world scientific literature. Each phase has a unique combination of (*a*) chemical formula and (*b*) modification. These are defined using the structure prototype, Pearson symbol, and space group. Each phase has an integer identifier called *phase id*. In the next three sections, each of the entry types (S-, P-, and C-) together with their interlinkage will be covered in details.

## 2.3   Physical Properties

The P-entries of PAULING FILE include the experimental and to a limited extent simulated data for a broad range of intrinsic physical properties of inorganic compounds in the solid, crystalline state. The considered physical properties belong to one of the following seven general domains: (*a*) electronic and electrical properties, (*b*) optical properties, (*c*) magnetic properties, (*d*) mechanical properties, (*e*) phase transitions, (*f*) superconductivity, and (*g*) thermal and thermodynamic properties. The taxonomy consists of three levels: the mentioned general domains, sub-domains, and the particular physical properties. For instance, the domain "electronic and electrical properties" contains the sub-domain "electron energy band structure," which in turn contains the "Fermi energy" property, etc. Currently there are about 100 sub-domains and nearly 2000 particular physical properties. The taxonomy was compiled by Fritz Hulliger (Swiss Federal Institute of Technology in Zurich, Switzerland), Roman Gladyshevskii (Ivan Franko National University of Lviv, Ukraine), and Karin Cenzual (University of Geneva, Switzerland). To a certain degree, it reflects the development of the solid-state physics during the last century.

The physical properties are stored in four different ways: numerical values, figure descriptions, property classes (such as ferromagnet, piezoelectric, etc.), and indications of existence of the particular data in a source publication, e.g., different spectra. The symbols for the most common physical properties have been standardized, mainly based on the *CRC Handbook of Chemistry and Physics* (Lide 1997–1998). The numerical values are stored in the published units and converted to the standard SI units. For certain properties at the atomic level, other units such as eV or $\mu B$ are used. Properties expressed with respect to a defined quantity of substance (per kg, per mole) are converted to per atom-gram. Each numerical property value is accompanied by information about the experimental conditions for the particular measurement. A great flexibility is provided via the links to the reference tables. Thanks to that, the new properties may be selected, and their symbols, units, and ranges of magnitude can be controlled.

All the data are taken from the primary literature. Each P-entry corresponds to a particular data source and can contain several numerical values, figure

descriptions, and keywords. For an investigation of a compound through a temperature- or pressure-induced structural phase transition, there will be two P-entries, for instance, one for the room-temperature modification and one for the low-temperature modification. By default, ferroelectric transitions are assumed to be accompanied by structural changes and will justify the creation of two P-entries, whereas magnetic, electric, or superconducting transitions are not. Data for the phases with a certain homogeneity range are grouped under a representative chemical formula. The actual composition for a particular measurement, when differing from the composition representing the P-entry, is specified among the parameters. As for the crystal structure part, there will be three database entries for a continuous solid solution between two ternary compounds: one for each ternary boundary compositions and a third one grouping samples containing four chemical elements. Some simulated data from the ab initio calculations are also included, in particular energy band structures, but focus is on experimentally measured data and values directly derived from measurements.

In addition to the physical properties (in the form of numerical values, figure descriptions, or keywords), and compulsory items such as the chemical formula, large amounts of information concerning the sample preparation and experimental conditions are stored. The following database fields may be present in a physical properties P-entry: (*a*) compound, such as chemical system, published chemical formula (investigated samples), representative standardized chemical formula, modification; (*b*) bibliographic data, such as reference, authors, language, title etc.; (*c*) preparation, such as starting materials and method of synthesis; (*d*) sample description, such as form, chemical analysis, stability with respect to temperature, pressure, and composition, elastic behavior, density, color, chemical reactivity; (*e*) crystallographic data, such as structure prototype, space group, and cell parameters.

The PAULING FILE data are checked for consistency using the original software package ESDD (evaluation, standardization, derived data), containing more than 100 different modules (Cenzual et al. 2000). The checking is carried out progressively level by level, also on the individual database fields. These include formatting of numerical values, units and symbols for physical properties, Hermann-Mauguin symbols, Pearson symbols, consistency of journal code, year and volume, pages for literature references, formatting of chemical formulae, usual order of magnitude, spelling, etc. Consistency checks within the individual datasets include atom coordinates, Wyckoff letters, site multiplicity, comparison of chemical elements in chemical system, chemical formula, and comparison of computed and published values. Further quantities for checking are cell volume, density, absorption coefficient, interplanar spacings, Pearson symbol, space group, cell parameters, refined composition, chemical formula, units, symbols for physical properties, Bravais lattice, diffraction conditions, site symmetry, anisotropic displacement parameters, and so on. Special checking of the crystallographic data includes comparison of the interatomic distances with the sum of the atomic radii, comparison of interatomic distances within chemical units, checks on charge balance, search for missed symmetry elements, and comparison with the type-defining entry (cell parameter ratios, atom coordinates). Consistency checks within

the entire database include comparison of densities, comparison of cell parameter ratios for isotypic compounds, check for compulsory data, check of database links, and so forth. Wherever possible, misprints detected in the original publications are corrected. Since editing mistakes can never be completely avoided, all modifications of the originally published data and interpretations of ambiguous data are stored in remarks. The ESDD software further computes the following parameters: at.% of the different elements, molar mass, refined composition and formula, computed density, interplanar spacings from functions of Bragg angle, equivalent isotropic displacement parameters, linear absorption coefficient, Miller indices referring to the published space group setting. It converts compositions expressed in wt.% to at.% and values expressed in various published units to standard units, including units per mole or wt.% to units per gram-atom, respecting the number of significant digits. As seen, an extreme care is taken to provide maximal quality of the stored data.

## 2.4   Crystalline Structures

Currently, PAULING FILE contains more than 350,000 crystalline structure S-entries. The minimal requirement for an S-entry in the PAULING FILE is a complete set of published cell parameters, assigned to a compound of well-defined composition. Whenever the published data are available, the crystallographic data also include atom coordinates, displacement parameters, and experimental diffraction lines and are accompanied by information concerning preparation, experimental conditions, characteristics of the sample, phase transitions, and dependence of the cell parameters on temperature, pressure, and composition. In order to give an approximate idea of the actual structure, a complete set of atom coordinates and site occupancies is proposed for S-entries where a prototype could be assigned. The crystallographic data are stored as published but also have been standardized according to the method proposed by Parthé and Gelato, using the program STRUCTURE TIDY (Gelato and Parthé 1987). When relevant, they are further adjusted so that the data for isotypic S-entries can be directly compared. Derived data include atomic environments of the individual atomic sites, based on the maximum gap method (Brunner and Schwarzenbach 1971; Daams et al. 1992), and the Niggli-reduced cell. The S-entries are checked for inconsistencies within the S-entry and by comparing different S-entries, using the program package ESDD mentioned in the previous section. For 5% of the S-entries, one or more misprints in the published crystallographic data are detected and corrected. Warnings concerning remaining short interatomic distances, deviations from the nominal composition, etc. are added in remarks. SI units are used everywhere, and the crystallographic terms follow the recommendations by the International Union of Crystallography.

Similarly to the physical properties, all the data are extracted from the primary literature. When available, supplementary materials deposited as CIF files or in the other formats are used as data source. Approximately 10% of the processed documents exist in an original version (e.g., Russian) and a translated version

(English); duplicates are avoided, and both references are stored. Crystallographic data, simulated by the ab initio calculations or optimized by the other methods, are only considered being confirmed by experimental observations. Distinct S-entries are created for all the complete refinements reported in a particular paper. For cell parameters without published atom coordinates, an S-entry is prepared for each chemical system and crystal structure. For example, for a continuous solid solution between two ternary compounds, there will be three S-entries: one for each ternary boundary composition and one for the quaternary system. The latter may contain a remark describing the composition dependence of the cell parameters. For the choice of the retrievable cell parameters, preference is given to values determined under ambient conditions.

All the S-entries are subdivided into different categories, according to the level of investigation: complete structure determined, coordinates of non-H atoms determined, cell parameters determined and prototype with fixed coordinates assigned, cell parameters determined and prototype assigned, only cell parameters determined, etc. In addition to the crystallographic data, large amounts of information concerning the sample preparation and experimental investigation are also included in the PAULING FILE. Basic data are stored as published (for rapid comparison with the original paper) and standardized (for efficient data checking and retrieval and for a homogeneous presentation). The following database fields may be present in a crystal structure S-entry: (*a*) classification, such as chemical system, published and standardized chemical formula, modification, colloquial name, structure prototype, Pearson symbol, space group number, Wyckoff sequence, mass per formula unit, computed density, level of structural investigation etc.; (*b*) bibliographic data, such as reference, authors, language, title; (*c*) published and standardized crystallographic data, including detailed information on the atom coordinates, etc. and transformation from published to standardized data; (*d*) Niggli-reduced cell, including transformation from published to Niggli-reduced cell; (*e*) isotropic, anisotropic displacement parameters; (*f*) published diffraction lines, Bragg angle or equivalent parameter, interplanar spacing, intensity, Miller indices, radiation, and remarks; (*g*) preparation, such as starting materials (purity, form), method of synthesis, etc.; (*h*) mineral name, and locality; (*i*) compound description, such as chemical analysis, stability, color, sample form, chemical reactivity, measured density, etc.; (*j*) determination of cell parameters and structure determination, such as sample, experimental method, radiation, and conditions; (*k*) figure descriptions, such as number in the original publication, title, parameters, and ranges; and, finally, (*l*) editor's or general remarks.

As said, each S-entry gets the structure prototype assigned. The structure prototype is a well-known concept in inorganic chemistry, where a large number of compounds often crystallize with very similar atom arrangements. The compilation Strukturbericht (Ewald and Hermann 1931) started already in the beginning for the twentieth century to classify crystal structures into prototypes, named by codes such as *A2*, *B2*, or *G1*. Though these notations are still in use, structure prototypes are nowadays generally referred to by the name of the compound for which this particular atom arrangement was first identified. The PAULING FILE uses a longer notation, which includes also the Pearson symbol (a lowercase

letter for the crystal system, an uppercase letter for the Bravais lattice, sum of multiplicities of all, fully or partially occupied atom sites) and the number of the space group from the *International Tables for Crystallography* (Hahn 1983). All datasets with the published atomic coordinates are classified into the structure prototypes. Each structure prototype is defined based on the first experimentally determined compound with the respective geometrical arrangement of atoms within the unit cell. More than 36,000 different structure prototypes have been identified and assigned to the S-entries. When not published, the editor also assigns the space group setting of the published cell parameters.

There exist an infinite number of ways to select the crystallographic data (cell parameters, space group setting, representative atomic coordinates) that define a crystal structure. The number remains high even when the basic rules recommended by the *International Tables for Crystallography* (Hahn 1983) are respected, due to the allowed symmetry operations such as permutations, origin shifts, etc. It follows that even identical or very similar atom arrangements may not be recognized as such. The classification of crystal structures into structure prototypes is largely facilitated by the use of standardized crystallographic data. The crystallographic data in the PAULING FILE are stored as published but also standardized. This second representation of the same data is such that compounds crystallizing with the same prototype (isotypic compounds) can be directly compared. It is prepared in a three-step procedure. First, the published data are checked for the presence of overlooked symmetry elements and, if found, converted into a space group of higher symmetry. Second, the resulting data are standardized with the program STRUCTURE TIDY. Third, the resulting data are compared with the standardized data of the type-defining database entry (Villars et al. 2018).

The atomic environments, also called coordination polyhedra, are defined for each S-entry using the method of Brunner and Schwarzenbach (Brunner 1971; Daams et al. 1992). One hundred different polyhedral types have been identified in the PAULING FILE data. In most structures, the coordination numbers vary from 1 to 22. It should be noted that this purely geometrical approach was developed for the intermetallic compounds and does not distinguish bonding types. As a consequence, the selected atomic environment may include both cations and anions. However, the method is simple to apply and useful in the majority of cases. Also, this approach offers an additional possibility to check the crystal structure data for geometrical correctness. The atomic environments can be used as the second independent structure classification. For instance, one can easily find geometrically similar prototypes. Notably, the PAULING FILE database supports geometrical restraint criteria for retrieval. That is, one may request information for the crystal structures containing, e.g., tetrahedra and octahedra.

## 2.5   Phase Diagrams and Distinct Phases

The phase diagram section of the PAULING FILE contains temperature-composition phase diagrams for binary systems, as well as the horizontal and vertical sections and liquidus or solidus projections for ternary, quaternary, and

other high-order systems. Both experimentally determined and calculated diagrams are processed. Primary literature is considered in the first priority. Also the diagrams from a few well-known compilations, such as the compendium of binary phase diagrams (Massalski et al. 1990) and the series of books on ternary phase diagrams (Petzow and Effenberg 1988–1995), have been included. All the diagrams have been converted to at.% and °C and redrawn in a single scale, so that the different reports for the same chemical systems can easily be compared. Single-phase fields are colored in blue and three-phase fields in yellow. The phases identified on the diagrams are named according to the PAULING FILE conventions, but also the original names are stored. Each phase diagram is linked to a C-entry, which usually contains the following database fields: (*a*) classification, such as chemical system and type of the diagram; (*b*) investigation, such as experimental or calculation technique, APDIC standardization; (*c*) bibliographic data, such as reference, authors, affiliation, language, and title; (*d*) original diagram details, such as figure number in the publication, borders, scales, sizes, etc.; (*e*) redrawn diagram details, such as concentration range, temperature, and conversion of concentration; and (*f*) list of the phases present on the diagram, standardized phase name, name used in the original publication, structure prototype assigned by the editor, structural information given in the original publication, and link to a representative crystal structure S-entry. For binary systems also the temperature and reaction type for the upper and (or) lower limit of existence of the phase are stored.

The physical property, crystal structure, and phase diagram entries are related via the distinct phases concept. At the database level, all three different types of data (P-, S-, and C-entries) are linked to the distinct phases table. To prepare this table, each chemical system has been evaluated. For example, the three major distinct phases reported for $TiO_2$ crystal (rutile, anatase, and brookite) are separated with respect to the temperature or pressure. Then the reported physical properties or crystalline structures of $TiO_2$ are associated only with the corresponding distinct phase. Finally, each C-entry (phase diagram) is formed by a particular set of the known distinct phases. Thus the unique interlinkage of data is achieved. A certain number of characteristics, attributed to the phases (compound classes, mineral names, etc.), are stored in the distinct phases table. Each distinct phase obtains a unique name containing a representative chemical formula, when necessary followed by a specification such as "ht," "rt," "cub," etc. There are the following special cases. *First*, the phases that crystallize in the same structure prototype, but are separated by a two-phase region in phase diagrams, are distinguished. The same is true for the temperature- or pressure-induced isostructural phase transitions where a discontinuity in the cell parameters is reported. *Second*, the structures with different degrees of ordering have in some cases been considered separately, in others not, depending on the possibility to assign unambiguously one or the other modification. Structure refinements considering, for instance, split atom positions are often grouped under the parent prototype. *Third*, the structure proposals, stated to be incorrect in the later literature, have been grouped under a single phase in agreement with the more recent reports. That is, e.g., an S-entry reporting a hexagonal cell may in such a case be grouped under an orthorhombic phase. *Fourth*,

the definition of a structure prototype used here suggests that a continuous solid solution may smoothly shift from one prototype to another. Refinements considering one or the other prototype are then grouped together. *Fifth*, the physical properties reported ignoring the crystal structure, and in principle referring to ambient conditions, are assigned to the "rt" modification, or, if the temperature dependence is not known, to the most commonly observed modification. *Sixth*, by default the paraelectric-ferroelectric phase transitions are assumed to be accompanied by a structural transition, and different phases are considered above and below the transition temperature. On the contrary, magnetic ordering is assumed not to modify the nuclear structure to a significant extent, therefore not leading to the new phases. Still there exist the chemical systems that are little explored, so that the reports in the literature are contradictory. The phase assignment becomes here very difficult, and the list of distinct phases will sometimes contain more phases than there exist in reality. It should be noted that there is a certain amount of subjectivity when assigning a phase identifier. Nevertheless, this approach represents a substantial advantage.

## 2.6   Applications

Thanks to the large amount of information stored in dozens of tables and hundreds of distinct database fields, the PAULING FILE offers almost unlimited possibilities for retrieval. It can of course be used for all kinds of trivial search, based on the chemical system, or literature data, but also much, much more. The conversion to standard units facilitates the search for properties within a particular numerical range, and the assignment of distinct phases plays an essential role, making it possible to combine searches on data stored in the three parts of the database: crystal structures, phase diagrams, and physical properties. The hundreds of interconnected database fields can be used to create different products. The PAULING FILE data are included in various printed products, as well as offline and online software, such as desktop catalogs, simulation environments, materials investigation toolkits, etc. Some of these products contain only structure data, others phase diagrams and crystallographic data, and others the three groups of data. Following the preference of the producers, some products contain only the published cell parameters, others only the standardized cell parameters, and yet others both published and standardized crystallographic data. Some of the products are limited to the PAULING FILE data, whereas others also contain data from other sources.

The Materials Platform for Data Science (MPDS) is a recent online infrastructure, presenting all the three parts of the PAULING FILE data. It contains nearly 70,000 phase diagrams, over 350,000 crystalline structures, and nearly 700,000 physical property entries. About 300,000 scientific publications in materials science, chemistry, and physics serve as source for these data. About 80% of the data can be requested remotely in a mass manner (via the so-called MPDS application programming interface) in a developer-friendly format, ready for any external modern data-intensive applications. Next sections give an overview of this

PAULING FILE implementation, its technical details and usage scenarios, as well as ongoing work in this field in general.

## 3    Materials Genome and Materials Platform for Data Science

### 3.1    Materials Genome Background

The concept of materials genome was taken from bioinformatics, referring to the Human Genome Project, publicly funded initiative started in 1990 and successfully ended in 2003 (Schmutz et al. 2004). In June 2011 the US President announced the multiagency Materials Genome Initiative to help US businesses and universities to discover, develop, and deploy new materials twice as fast, at a fraction of the cost. In 2012, Materials Genome Initiative commitments include $12 million of research at the Department of Energy and $17 million in materials research at the Department of Defense. To the end of 2014 several dozens of universities were participating. The oldest and, probably, the most recognized participant is the Materials Project, an undertaking of the groups of Gerbrand Ceder and Kristin Persson (Lawrence Berkeley National Laboratory). The Materials Project team had identified hundreds of new compounds, several of which now function as lithium battery electrode materials. The software toolkit for materials design, development of the Materials Project, simplifying routine computational tasks, is actively used by about 100 scientists in different organizations around the world, and approximately every second user contributes in the open-source code. Additionally, there exists an open database, prepared using this toolkit (approx. 100,000 compounds). Importantly, the Materials Project team on a half consists of the experimental scientists, who deal with lab synthesis. This leads to an incredibly strong collaboration of theory, modeling, software development, and experiment.

The first European counterpart of the Materials Project was the NoMaD Project, started in 2013 from the collaboration of Fritz Haber Institute (FHI) of the Max Planck Society, Berlin, and Humboldt University of Berlin (HU) with the aim to create an international ab initio materials science data repository. In the end of 2014, the first version of NoMaD user interface was publicly launched, and in 2015 NoMaD Project was successfully funded by European Union's infrastructure call for Centers of Excellence (CoE) in computational sciences. As of 2017, the NoMaD data repository contains more than three million ab initio simulation files (more than 10 Tb disk space on estimate), contributed by the community and taken from the other repositories. In total, more than ten data formats for all the major well-known quantum simulation packages are supported. Based on these data volumes, the online materials science encyclopedia and the software analytics toolkit are publicly released. A possible disadvantage of NoMaD is the focus on the community's data centralization, which is currently not very well accepted in the materials science community.

Speaking about the publicly funded projects of materials genome, an ongoing commercial activity should also be mentioned. The notable product with the long

history (since 1998) is MedeA scientific software environment (Christensen et al. 2017), which presents one of the most advanced and sophisticated simulation desktop toolkits. The MedeA employs computational workflows and allows to manage high-throughput database-driven simulations. Also there are two prominent materials informatics startups in the San Francisco Bay Area, USA: Exabyte and Citrine. The focus of the *Exabyte* software platform is the high-throughput ab initio simulations of the materials, performed in the cloud, i.e., at the commodity hardware cluster infrastructure, rented at one of the public vendors. As a simulation engine, the well-known VASP package is employed. In two recent years the team was able to perform a comparative analysis of about a thousand different materials by utilizing extensive on-demand scalability of the developed cloud platform (Bazhirov et al. 2017). The total costs of the runs ended up not exceeding a few thousand US dollars. The *Citrine*'s online data platform is called Citrination (O'Mara et al. 2016). It was launched in 2015 and now houses over several millions of data points. So far the platform has received various contributions from about 2000 different institutions worldwide, including universities, government laboratories, and companies. The disposed data are completely free and opened. Citrine itself contributes to its platform, searching and disposing the datasets from the opened online sources. The platform provides a free mass access interface for all its data. Citrine also claims to develop the artificial intelligence-based tools that enable new insights from the collected materials data.

It is seen that all the abovementioned initiatives have one main feature in common, namely, they build their own software infrastructures to process information efficiently and tackle the challenge of materials big data. They also develop the novel data-intensive analysis methods. The development of such tooling is mostly conducted within the *open-source* paradigm. This means, the program code of the complex tools is provided for free, and anyone can adopt it for own aims. This has a strong rationale. After a certain complexity threshold, the software product becomes practically unusable, because only a very limited group of professionals are able to deal with it. Often such people are not motivated by money, but they can be attracted by chances and challenges of the possible technological breakthrough. Thus, open-sourcing the parts of the code is great advertising, which allows to attract such talents that could hardly financially be attracted. If the code is popular enough for the outside contributions, a force multiplier is created that helps to get more work done faster and cheaper. More contributors mean more use cases being explored and, finally, the more robust software. Importantly, the user community must be grown around the open-source tooling. Normally such community is fairly amorphous and requires guidance and patronage. The expenses for the development are well covered by the talented contributions from outside, reputation and acknowledgment, which, in turn, can be converted to the other means of profit. It is also planned to open-source the certain parts of the software of Materials Platform for Data Science. And a part of the PAULING FILE data is now already opened online under the Creative Commons Attribution 4.0 International license. These are (*a*) all entries found by keywords "cell parameters – temperature diagrams" and "cell parameters – pressure diagrams," (*b*) all data for compounds containing both Ag and K, and (*c*)

all data for binary compounds of oxygen. In total, (*a–c*) present quite a rich dataset, suitable for educational data-mining purposes.

## 3.2    Development of the PAULING FILE Materials Infrastructure

As of today, materials informatics is a collection of recipes taken from computer science and adopted for the modern materials science. The main difficulty is purely technical from an academic point of view – how to handle materials big data efficiently. Here some recipes are discussed. All they were considered while development and maintenance of the PAULING FILE online retrieval software – MPDS (Materials Platform for Data Science).

One of the possible ways to tackle complexity is the unified modeling language (UML), general-purpose, developmental, modeling language for software engineering (see e.g., Miles and Hamilton 2008). From a bird's eye view, UML is a convention of drawing concepts in a human-understandable manner. Although it cannot be interpreted by a computer, UML provides formal description of the problematic field, which is then much easier to encode in a computer programming language. Another UML advantage is its standardization and high popularity in the field of software engineering (suited for collaborative work). The concepts in UML must be related with the defined relationship types.

The PAULING FILE concepts as implemented in the MPDS platform are represented in Fig. 1, with the very short UML legend at the bottom. Notably, UML provides a clear formal way to understand, how a certain problem domain is organized. Namely, it is seen at Fig. 1 that (*a*) any data-mining tool (e.g., visualization) is based on the MPDS PAULING FILE data; (*b*) data are subdivided into three parts: crystalline structure part (S), physical properties part (P), and phase diagrams part (C); (*c*) each part is represented by entries; and (*d*) each entry is concerned with the relevant phase and scientific publication. Also, there are users with the different data access permissions. Thus, the UML presents the important guideline for the further development.

The MPDS is an online software, working according to a client-server architecture. There are many advantages of the online products over the offline products: absence of installation, cross platform operation, no special requirements to the client PC, transparent updates, enhanced security and reliability, and more. With the ubiquitous penetration of the Internet and the wide availability of the server resources, the online model becomes clearly preferable. The details are presented in Fig. 2.

Generally, the World Wide Web is based on the idea of interconnection. Indeed, in a modern environment no isolated software per se makes sense, and each application communicates with the others. In order to provide fast and efficient interaction experience and not to develop a new access interface for each case, the application programming interfaces (APIs) are commonly used. The API regulates communications between any kind of the software (be it a complex of data-mining programs, simulation platform, or any other big data consumer). The main idea is
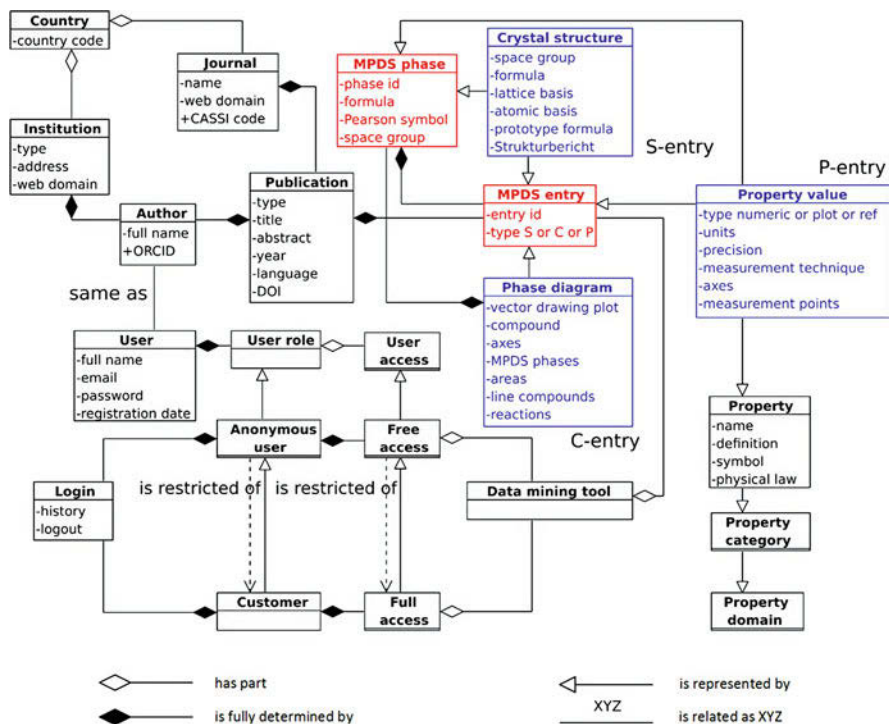
**Fig. 1** UML diagram for PAULING FILE concepts, as implemented in the MPDS

that the functionalities are collected in a single place and exposed (encapsulated) via the online API. The APIs in the online medium normally adhere to the principle of representational state transfer or REST (Fielding 2000). The REST presents guiding constraints for client-server software architecture and could be called as meta-API. It is also employed for the MPDS API (see Fig. 2), which presents all the PAULING FILE data in a developer-friendly, machine-readable way, using the opened formats, such as CIF, JSON, and MIF (see below). Importantly, the API is not only how the clients communicate with the server. In a wider sense, online API is a software architecture, declaring the way of all the communications. This way the audience is not bound to the existing human-oriented graphical user interfaces (irrespective of their convenience) and able to use the service provider maximally efficient for their aims.

Nevertheless, much attention at the MPDS is given to its graphical user interface. The search input field allows to type different combinations of basic search terms seamlessly at once, so that they are correctly recognized and the matching entries are shown. The algorithm, responsible for treatment of such seamless searches, works as follows. First, the stop-words ("the," "and," "about," etc.) are filtered. Then, the words are checked to belong to the one-term categories: chemical formulae, chemical elements, and crystalline lattices. These categories cannot contain more
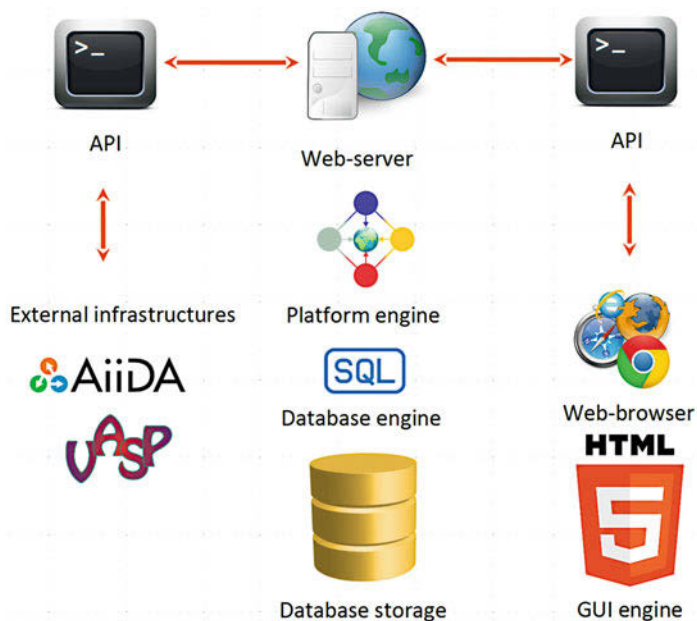
**Fig. 2** Client-server architecture of the MPDS

than a single word. The found words are interpreted and excluded. Then, the remaining words are checked to belong to the materials classes and physical property categories. The matches of the longest combinations of words are checked in the beginning, and then the shorter combinations are checked. Unrecognized words are marked as "ignored."

From the very beginning of MPDS development performance of all the sub-systems was a cornerstone. This is especially important for the centralized online software, when the server is a single point of failure. Generally, for any online project it is always recommended to focus on the operation speed. It should be noted that the software designed for 5 simultaneous users differs considerably from that designed for 50 or 500 simultaneous users. There are two approaches to scaling the performance: vertical and horizontal (Michael et al. 2007). *Vertical scaling* means the more powerful server is deployed with the increased loads, but the software architecture remains the same. This solution is a quick and efficient, although limited. First, the more powerful, the more expensive is the server, and the dependence is not linear. Second, there is a limit of computational capacity per a single server. And the database performance is normally saturated much lower this limit. So for a database after a certain tolerance vertical scaling makes no sense. The solution in this situation is the *horizontal scaling*, when the software architecture is changed in such a way to distribute the increased loads evenly among the server cluster, consisting of inexpensive commodity hardware. The idea is that the new replicas with the MPDS software are added to the server cluster by

simple quick copying. A new replica joins the server cluster and starts accepting the incoming HTTP requests. Furthermore, if a replica fails by whatever reason, it can be automatically replaced by a new one. This way the server cluster stays highly available, and even the huge loads cannot influence the performance, since more cheap replicas could be added. However while the server price remains under control, the development is complex and requires time.

## 3.3 Storage and Exchange of Materials Data

Similar to any other scientific data, materials data are commonly stored in the files and databases. The files usually present raw data (simulation or measurement logs, compressed archives, etc.) Reusable processed data are normally stored in the databases. It should be noted although that such subdivision is very conditional, moreover a file could be a database and vice versa (e.g., HDF5 container, SQLite, etc.). A database model is a type of data model that defines the logical structure of a database and fundamentally determines in which manner data can be stored, organized, and manipulated.

The MPDS employs several database models, in particular, relational and semantic graph model. Additionally, the document-oriented model is used as an intermediate step while preparing data. In the document-oriented model each PAUL-ING FILE entry is treated as a *document*. Such documents may have varying number of properties and are all stored in a single giant database table. All the document properties are searchable. The model is implemented using the PostgreSQL DBMS. Being very simple, this model allows unique flexibility (single table, arbitrary data queries). The cost of such flexibility is very low access speed. Being inappropriate for production, it suits very well for development. Being robust and mature, the relational model is a core of the MPDS production system. It is also supported by the PostgreSQL DBMS, in addition to the intermediate document-oriented model. The standard data manipulation language within the relational model is called structured query language (SQL) and based upon the relational algebra. Eventually, the single table from the document-oriented model is taken, refined, and split into many simpler tables related to each other. This process is called normalization (Teorey et al. 2005). This way much greater access speed is achieved. The refined table structure is however tightly bound to the chosen usage scenario and does not provide extra flexibility.

The disadvantage of the relational model is that the data must obey the strict rules, defined as the database schema. Normally, these rules do not imply an existence of other terms at the database level outside an application business logic. However, the expert systems and artificial intelligence applications must act in an opened world, making inferences and determining new facts, basing on the newly collected information. Their databases must be able to include the new terms and to follow the new logic dynamically and therefore do not fit into the traditional relational paradigm. Usually, such databases employ the so-called semantic Web approach (DuCharme 2013). The data in such model are represented as a *graph*. Such graph

is comprised by a set of statements in the form "subject-property-object," called triples. Using the semantic graph database model, the unprecedentedly flexible and expressive queries on top of the knowledge graph become possible. The MPDS currently uses the semantic graphs only indirectly, complementary to the relational model; however much wider adoption of this model is planned. As a back end for the semantic graph model, the Virtuoso DBMS is employed.

As mentioned, the development of the single data exchange format in materials science is an extremely complex task. Nowadays computer science suggests a very convenient paradigm, when a specific information container is accompanied with the rules definition, i.e., an automatic validation tooling. There are various examples following this idea: SQL and schemata for the relational databases, A-Box and T-Box for the semantic graphs, XML and XSD for the machine-readable data transfer and markup, JSON and schemata for the human-readable data transfer, etc. In this respect, two successful achievements in developing the common exchange format in materials science should be mentioned: Crystallographic Information File (Hall et al. 1991) and Materials and Physical Information File (Michel and Meredig 2016). Both of them are supported at the MPDS and partially at the other mentioned materials informatics infrastructures.

The Crystallographic Information File (CIF) was established in the 1990s by the International Union of Crystallography (IUCR). CIF is based on a text container called STAR (Self-Defining Text Archive and Retrieval), where the physical properties, obtained, e.g., as a result of X-ray diffraction or theoretical modeling, are labeled by the standard tags. The standard tags determine the parameters of the unit cell, symmetry, atomic positions, relevant scientific publication metadata, etc. These tags are defined in the external CIF dictionaries (cf. XSD schemata for the XML documents), so it is possible to validate a CIF file against a CIF dictionary and even to infer the new physical properties from those available. The difference is that the CIF format allows the arbitrary tags. They are ignored by CIF parser but later can become the part of standard CIF dictionaries, according to IUCR. Furthermore, CIF format supports the relational data model, so one can refer to the specific atom in the crystalline structure by its identifier. The drawback is the absence of a convenient multilevel hierarchy support, so here the STAR container concedes to XML. CIF format is used for the online crystal structure visualization at the MPDS, and no other proprietary or self-made formats are employed. Only the Web browser is needed for visualization, and no plugins, applets, or other software is required. Normally, structure rendering is done on the GPU (i.e., graphical card). However, if the GPU is outdated or not available, rendering is done on the CPU (central processor). In this case the quality of rendering is reduced. The total size of the code served online for the rendering is only about 150 Kb, and after the code is loaded, no further Internet connection is required. The CIF visualization at the MPDS is based on the open-source technologies.

The JSON format is simpler, more flexible, and more permissive. Historically, it is much more common for the software development than a narrow-purpose CIF. The container of the Materials and Physical Information File (MIF), introduced by a Citrine startup, is built on top of JSON, taking all its advantages. JSON

also provides its own schema approach, used for validation, documentation, and interaction control, i.e., a contract for the JSON data required by a given application, and how that data can be modified. JSON schemata for all types of the PAULING FILE data, including MIF specifications, were developed in 2016 and now are publicly available. The JSON format, including MIF, is the most widely adopted within the mentioned materials infrastructures.

## 3.4   Computer-Assisted Data Analysis

Undoubtedly plotting and visualizations are extremely helpful for data analysis. One of the interesting plotting features of the MPDS is the semantic graphs of terms. Generally, all the MPDS data are a giant semantic graph of the structured knowledge in materials science, accumulated by the humankind. Unfortunately, no human being may observe this graph as a whole. The online interactive visualizations attempt to show only tiny portions of a giant MPDS semantic graph, related to the particular user's input, in a very simplified form. Fortunately, modern semantic technologies (DuCharme 2013) are able to comprehend a giant MPDS graph all at once. This is the planned direction of the MPDS platform development in future. Another type of visualizations, the dynamically rendered phase diagrams (C-entries), displayed online at the MPDS, are fully digitized, programmatically drawn plots. The rendering engine works in all the modern Web browsers, requires no plugins, and is based on the open-source Web technologies. The phases at the phase diagram are associated with the parametric equations in the form $x = x(t)$, $y = y(t)$, $0 <= t <= 1$, where $x$ stands for the composition and $y$ for the temperature.

Recently the MPDS has launched the first version of its machine-learning predictions. To demonstrate some practical usage scenarios of the materials data-mining using MPDS API, a relatively unsophisticated yet powerful predictive machine-learning algorithm, the decision tree regression, was chosen. A decision tree is a statistical model, which describes the data going from the observations about some item (e.g., a crystalline structure) to the conclusions about the item's target value (e.g., a corresponding physical property). The MPDS data contain crystal structures with the corresponding physical properties, so it is feasible to train a model on this dataset. The following physical properties were chosen: isothermal bulk modulus, enthalpy of formation, heat capacity at constant pressure, and melting temperature. Multiple decision trees were built by repeatedly resampling training data with replacement and voting for the trees yielding a consensus prediction. This algorithm is known as a "*random forest*" regressor. The "random forest" is a statistical estimator that fits a number of classifying decision trees on various subsamples of the dataset and use averaging to improve the predictive accuracy (Breiman 2001). Its presently used state-of-the-art open-source implementation is very efficient and takes seconds to train a model from MPDS data on an average desktop PC (McKinney 2010).

The evaluation process was repeated at least 30 times to achieve a statistical reliability. The results of a randomly chosen evaluation process are shown in Fig. 3.
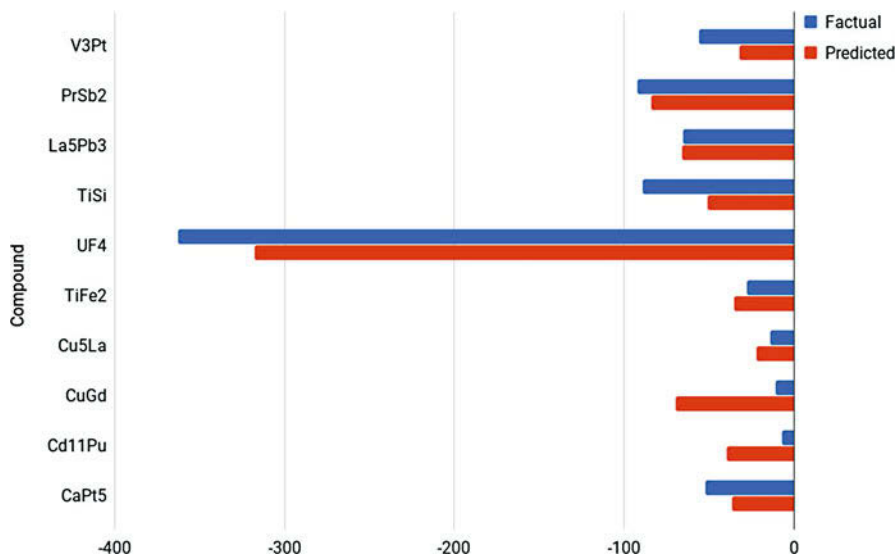
**Fig. 3** Formation enthalpy, kJ g-at.$^{-1}$: *predicted* versus *factual* values; mean absolute error is 42 kJ g-at.$^{-1}$, comparable with the results of the ab initio simulations

The prediction quality is acceptable and on average may even compete with the ab initio simulation results. The difference is that the simulation normally requires hours or days of computation time, whereas the machine-learning model yields the results in milliseconds on the same hardware. Another difference is that the ab initio simulations in practice require careful fine-tuning of the method, whereas the chosen method of machine learning is a black box, where no initial setup is needed. The disadvantage of the machine-learning model is that no physical meaning of predictions is implied. The underlying complex physical phenomena, as well as the lack of training data, may lead to the poor prediction quality. The size of the training dataset should not be necessarily huge, but there is some minimal threshold. For example, predictions using the smaller dataset of the open MPDS data demonstrate worse quality. Here it is important to note that nowadays more accurate machine-learning techniques exist, such as deep learning neural networks, and the quality of predictions may be further increased.

In the search for the new technological opportunities in the materials data mining, the end point of the route is the artificial intelligence techniques. This is an understandable fact, since the more sophisticated technology, the more similar its intellectual output to the human's one. Historically, there are two approaches to the construction of the intelligent agents (Jones 2008): statistic ("bottom-up" or connectionist) and deterministic ("top-down" or symbolic). The statistic approach is known as "machine learning," i.e., the machine is expected to discover the world on its own, as the humans do. The deterministic approach is known as "inference engine," i.e., the facts are prepared in advance and considered logically. This

approach will be shortly reviewed below. The so-called *ontologies* are employed. The language for ontology expression is called OWL (ontology Web language) and based on the logical calculus, used in mathematics, philosophy, linguistics, and computer science (Baader et al. 2007). Applied to materials science, the ontologies are almost unknown, although this term appears in the publications last two decades from time to time. Conceptually, the ontologies and databases look similar. A database consists of the schema and data (i.e., arrangement of tables plus their content). An ontology consists of the axioms and facts (called T-Box and A-Box). There is however an important difference compared to databases. A database schema acts as constraints on structure of data, whereas ontology axioms act as implications or inference rules. The logical expressiveness of the ontology axioms can be much higher (richer) than any database schema. Currently the schema of the MPDS platform database is quite carefully tuned to meet the designed usage scenarios. On the other hand, this is also a potential limitation. Being highly optimized to satisfy online searches, the MPDS platform database does not however perform very well for nonstandard complex queries and hierarchical data manipulations. In this sense, the ontologies fit the MPDS data model very well. Nonetheless, currently the most important drawback of the ontologies compared to the databases is that the performance cannot be unfortunately even theoretically guaranteed, as the OWL logical calculus is extremely complex mathematically. Apparently, all the successful production deployments of the ontologies somehow overpass this limitation (Cuel and Young 2015). Notably, the ontologies per se do not present much practical sense. In fact, they are just logical constructions attached to the data. Only the whole set of accessory technologies and implementations (called the *semantic web stack*) is able potentially to shed more light to the laws of nature, enclosed in the materials data.

## 3.5    Data-Centric Observations

Following Dmitri Mendeleev and Lothar Meyer, who observed periodical patterns in the properties of the chemical elements in the 1860s, one may try to find similar patterns in the whole range of the known materials and make a step toward hypothetical periodic table of materials. Based on the observations across the PAULING FILE database, the 12 empirical principles were formulated (Villars 1995; Villars et al. 2008; Villars and Iwata 2013). These principles can be called the cornerstones of nature, and all they can be explained using the modern electronic structure theory. They define (*a*) the compound formation, (*b*) the ordering of chemical elements within a structure prototype, and (*c*) the linkage between the position of a chemical element in the periodic table and its occupied sites in the structure prototype.

*First*, compound formation: the atomic size, electrochemical, valence-electron, and cohesion energy factors are governing compound formation. For example, one can observe that about 30% of all chemical element combinations form no compounds within the binary, ternary, and quaternary systems.

*Second,* correlation of the number of chemical elements and atomic environment type: the maximal diversity of the atomic environments is achieved within the binary and ternary inorganic compounds. The quaternary, quinary, and other higher-order compounds strongly prefer the atomic environment types with the low coordination number. One may also call this observation as a surprising reduction of geometrical diversity with the increasing element count.

*Third*, active concentration ranges: there is the systematic occurrence of daltonide inorganic phases for binary, ternary, and quaternary inorganic systems within the certain rather limited (active) concentration ranges.

*Fourth*, stoichiometric ratio condition: there are highly preferred stoichiometric ratios for the vast majority of daltonide compounds.

*Fifth*, compound simplicity: the vast majority of the inorganic compounds have on average only ten atoms per unit cell, thus showing only three or fewer atomic environment types within the crystal structures. One may note here that the nature indeed prefers simplicity.

*Sixth*, compound symmetry: 10% of the space groups cover nearly 70% of the inorganic compounds. The most frequent 11 space groups are 12, 62, 63, 139, 166, 191, 194, 216, 221, 225, and 227. As seen, the high symmetry is preferred.

*Seventh*, atomic environments arrangement: 18 out of about 100 possible atomic environment types are highly preferred and were found for 90% of all the PAULING FILE compounds. In particular, the most frequent types of polyhedra are tetrahedron, octahedron, cube, tri-capped trigonal prism, four-capped trigonal prism, icosahedron, cubooctahedron, bi-capped pentagonal pyramid, and anti-cubooctahedron.

*Eighth*, chemical element ordering tendency: only about 30 structure prototypes have more than 1000 representatives, and the 1000 most populous prototypes and their representatives cover the majority of the crystalline structures of the PAULING FILE.

*Ninth*, correlation of the structure prototype and the periodic system: the vast majority of the crystalline structures show a very strict regularity between the position of the chemical element in the periodic system and its Wyckoff position occupation. This is confirmed on an example of the 1000 most populous prototypes.

*Tenth*, linking of structure and stability: the atomic size, electrochemical, valence-electron, and atomic number factors determine the crystalline structures of the intermetallic compounds. Again, this is applicable to binary, ternary, and quaternary systems. One may reveal clear patterns for, e.g., former versus non-former systems, iso-stoichiometric structure stability maps, and complete solid solubility between binary compounds within the same prototype.

*Eleventh*, generalized atomic environment type stability: using the periodic number (from Lothar Meyer's periodic table), one may subdivide different atomic environment types into distinct stability domains. It was found that the chemical elements with the periodic number more than 54 control the atomic environment types, independently of whether they act as the central or coordinating atoms. Thus, there exists a clear separation between the possible and impossible atomic environment types. Interestingly, the diversity of atomic environment types is very much reduced for quaternaries, as compared to binaries and ternaries.

*Twelfth*, complete solid solution stability: the atomic size, electrochemical, and valence-electron factors control solid solubility. For example, in a ternary system, where two of the binary boundary systems have the same structure prototype, one may predict whether a complete or limited solid solution is formed. Similarly, for a given chemical element, either a limited or extended solid solubility can be predicted.

As seen, all intrinsic physical properties of a single-phase inorganic solid are strongly linked to its crystal structure, emphasizing the importance of the crystal structure classification. And the 12 principles outlined above can only be discovered by the examination of a large amount of critically evaluated experimentally determined data. They ultimately lead to the restraints, which are a requirement for the development of a practicable and trustworthy computational materials design approach.

## 3.6   Applications

A trustworthy linkage between the published experimental inorganic solids and the high-throughput DFT calculations opens the new perspectives of the database-driven, data-intensive research and discovery of the new materials. At the moment two such high-throughput computational initiatives in China and Switzerland are starting to employ the MPDS API with the PAULING FILE data.

The majority of nowadays' practical materials science problems (clean energy sources, energy storage, superconductivity, etc.) are concerned with the big amounts of the complex knowledge to be assimilated, if not by the gifted humans then by the artificial intelligence agents. With the progress of the neuroscience and medicine, shedding the light on the nature of the human brain, new powerful computer science techniques emerge and later find their application in materials science. However, development of the complex tools in a closed (proprietary) environment is extremely difficult and even inefficient. This is why an academic tradition of openness and free exchange of the ideas realizes in an open-source strategy, showcasing the economically advantageous altruism.

## 4   Summary

The PAULING FILE, a unique and probably the oldest effort on organizing the materials data, was reviewed. Starting from the foundations of materials science, such as the taxonomy of physical properties, concepts of the structure prototypes, phases, phase diagrams, etc., it spans to the modern data-intensive science, employing the novel data storage and analysis techniques and providing high-quality materials insights, as implemented in the MPDS online platform. This is important while the efficient and quick knowledge exchange in materials science is obstructed, ahead of today's big data challenge. The other efforts and their contributions to the materials informatics are also noted. An importance of

fostering the emerging ecosystem of today's materials informatics is emphasized. It is based on the standards of JSON, MIF, and CIF, supporting the databases, virtual laboratories, connected with the online REST APIs, which tend to become the software infrastructure backbones. A conceptual UML modeling of the PAULING FILE at the highly abstraction level was showcased. Three types of the PAULING FILE scientific data are highly interlinked, therefore, inseparable within the MPDS online platform. The rich data structure assumes combination of storage approaches, in order to benefit of their advantages and mitigate deficiencies at the same time. Exploring the large amounts of different materials data, a holistic view on inorganic substances may be presented. Coupled with the high-throughput ab initio simulations, it can provide a key to the discovery of materials genome, playing a role of periodic table for entire set of materials. There are certain hopes for the artificial intelligence techniques, automatically generating the new materials discovery ideas, and, citing Linus Pauling, the best way to have a good idea is to have lots of ideas.

# References

Baader F, Horrocks I, Sattler U (2007) Description logics, Chapter 3. In: Handbook of knowledge representation. Elsevier, Amsterdam

Bazhirov T, Mohammadi M, Ding K, Barabash S (2017) Large-scale high-throughput computer-aided discovery of advanced materials using cloud computing. APS March Meeting abstract id. C1.007 https://ui.adsabs.harvard.edu/abs/2017APS..MAR.C1007B/abstract

Bornmann L, Mutz R (2015) Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J Assoc Inf Sci Technol 66(11):2215–2222

Breiman (2001) Random forests. Mach Learn 45:5

Brunner GO, Schwarzenbach D (1971) Zur Abgrenzung der Koordinationssphäre und Ermittlung der Koordinationszahl in Kristallstrukturen. Z Kristallogr 133:127–133

Cenzual K, Berndt M, Brandenburg K, Luong V, Flack E, Villars P (2000) ESDD software package, copyright: Japan Science and Technology Corporation, updates by O. Shcherban. Structure-Properties Company, Lviv

Christensen M et al (2017) Software platforms for electronic/atomistic/mesoscopic modeling: status and perspectives. Integr Mater Manuf Innov 6(1):92

Cuel R, Young R (eds) (2015) Formal ontologies meet industry. In: 7th international workshop FOMI 2015 Berlin Germany, Proceedings, Springer International, Switzerland

Daams JLC, van Vucht IHN, Villars P (1992) Atomic-environment classification of the cubic "intermetallic" structure types. J Alloys Compd 182:1–33

DuCharme B (2013) Learning SPARQL, 2nd edn. O'Reilly Media, Sebastopol, CA

Ewald PP, Hermann C (eds) (1931) Strukturbericht. Akad. Verlagsgesellschaft M.B.H, Leipzig

Fielding R (2000) Architectural styles and the design of network-based software architectures. Doctoral dissertation, University of California, Irvine

Gasteiger J, Engel T (2003) Chemoinformatics: a textbook. Wiley, Weinheim

Gelato LM, Parthé E (1987) STRUCTURE TIDY- a computer program to standardize crystal structure data. J Appl Crystallogr 20:139–143

Ghiringhelli LM, Vybiral J, Ahmetcik E, Ouyan R, Levchenko SV, Draxl C, Scheffler M (2017) Learning physical descriptors for materials science by compressed sensing. New J Phys 19:023017

Hahn T (ed) (1983) International tables for crystallography, vol A. In: Reidel D (ed) Springer, Dordrecht

Hall SR, Allen FH, Brown ID (1991) The Crystallographic Information File (CIF): a new standard archive file for crystallography. Acta Crystallogr A47(6):655–685

Jones MT (2008) Artificial intelligence: a systems approach. Jones & Bartlett Learning, Sudbury

Kong C, Villars P, Iwata S, Rajan K (2012) Mapping the materials gene for binary intermetallic compounds–a visualization schema for crystallographic databases. Comput Sci Discov 5:1

Lide DR, editor-in-chief (1997–1998), CRC Handbook of chemistry and physics, 78th edition, CRC Press, Boca Raton.

Massalski TB, Okamoto H, Subramanian PR, Kacprzak L (eds) (1990) Binary alloy phase diagrams, 2nd edn. ASM International, Materials Park

McKinney W (2010) Data structures for statistical computing in Python. In: Proceedings of the 9-th Python in science conference, p 51

Michael M, Moreira J, Shiloach D, Wisniewski R (2007) Scale-up x scale-out: a case study using Nutch/Lucene. In: 2007 IEEE international parallel and distributed processing symposium, p 330

Michel K, Meredig B (2016) Beyond bulk single crystals: a data format for all materials structure–property–processing relationships. MRS Bull 41(8):617–623

Miles R, Hamilton R (2008) Learning UML 2.0: a pragmatic introduction to UML. O'Reilly Media

Murray-Rust P (2013) Personal communications

O'Mara J, Meredig B, Michel K (2016) Materials data infrastructure: a case study of the Citrination platform to examine data import, storage, and access. J Miner, Met Mater Soc 68:2031

Obama B (2011) Materials genome initiative of the US Government. https://obamawhitehouse.archives.gov/mgi

Petzow G, Effenberg G (1988–1995) Ternary alloys: a comprehensive compendium of evaluated constitutional data and phase diagrams, 15 vols. Wiley-VCH, Weinheim

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci 111:218–230

Rajan K (2015) Materials informatics. Mater Today 15:470

Schmutz J, Wheeler J et al (2004) Quality assessment of the human genome sequence. Nature 429:365

Teorey T, Lightstone S, Nadeau T, Jagadish H (2005) Database modeling and design, 4th edn. Elsevier, Amsterdam

Villars P (1995) In: Westbrook JH, Fleischer RL (eds) Intermetallic compounds, principles and practice, vol 1. Wiley, New York, pp 227–275

Villars P, Cenzual K, Daams J, Chen Y, Iwata S (2004) Data-driven atomic environment prediction for binaries using the Mendeleev number: Part 1. Composition AB. J Alloys Compd 367(1–2): 167–175. https://doi.org/10.1016/j.jallcom.2003.08.060

Villars P, Daams J, Shikata Y, Chen Y, Iwata S (2008) Data-driven generalized atomic environment prediction for binary and multinary inorganic compounds using the periodic number. Chem Met Alloys 1:210–226

Villars P, Iwata S (2013) PAULING FILE verifies/reveals 12 principles in materials science supporting four cornerstones given by nature. Chem Met Alloys 6:81–108

Villars P, Cenzual K, Gladyshevskii R, Iwata S (2018) PAULING FILE – towards a holistic view. In: Materials informatics. Wiley

Xu Y, Yamazaki M, Villars P (2011) Inorganic materials database for exploring the nature of material. Jpn J Appl Phys 50:11S

# Crystallography Open Database (COD)

# 85

Saulius Gražulis, Andrius Merkys, and Antanas Vaitkus

## Contents

### Abstract

The Crystallography Open Database (COD, http://crystallography.net/) is as of the time of writing the largest open-access collection of mineral, metal organic, organometallic, and small organic crystal structures, excluding biomolecules that are stored separately in the Protein Data Bank (http://wwpdb.org/). Unlike other existing chemical crystal structure databases, the COD is fully open – all its structures may be downloaded, used, and re-disseminated without restriction, along with the results derived from them. Currently, the COD contains >385,000 records and is growing constantly, encompassing most structures published in peer-reviewed academic press and donations by individual researchers. This

S. Gražulis (✉) · A. Merkys · A. Vaitkus
Department of Protein-DNA Interactions, Vilnius University Institute of Biotechnology, Vilnius, Lithuania
e-mail: grazulis@ibt.lt

article describes how data are organized in the COD and how the database can be queried, downloaded, and processed for various purposes.

# 1    Introduction

X-ray crystallography is an extremely powerful method for determining inner structure of the condensed matter. Soon after the discovery of X-rays (Röntgen 1896) and the first records of their diffraction on crystalline samples (Friedrich et al. 1912a, b), the number of structures determined by this technique started to grow. An explanation of the X-ray scattering using first principles (Bragg and Bragg 1913; Bragg 1913) allowed determination of structural models for a vast variety of solid materials in a uniform way, from simple inorganics to very large biomolecules. As more and more crystal structures were appearing, it became evident that the numbers (such as crystal unit cell parameters, atomic coordinates) in their descriptions, made uniform by the availability of the common scattering theory, possess a great value themselves and efforts to collect them systematically were started. The first collections were in paper form (Hermann and Ewald 1931; IUCr 2017c; Kitaigorodsky 1955), and numeric data were accompanying crystallographic publications in journals dedicated for this field from the very first publications (for instance, in the Acta Crystallographica journal started by the IUCr in 1948 (Clews and Cochran 1948).

Growing availability and power of electronic computers allowed crystallographers to use them for structure determination and prompted the idea that crystal structure data can also be handled automatically (Brown and McMahon 2002). A first dedicated crystallographic database, the CSD, was established by the CCDC in 1965 (Groom and Allen 2014) to collect structures of small organic molecules and embraced computer-assisted methods for information storage and retrieval (Allen et al. 1979). Data about inorganic crystals (Kaduk 2002), alloys (White et al. 2002) and powder diffraction data (Kabekkodu et al. 2002) were historically kept in separate archives. Today, we have a whole range of databases, differing by their scope, size and licensing model, covering various aspects of crystallographic data (Table 1).

As seen in from the Table 1, various licensing models were employed to support operations of the databases. About a third of all resources, and some of the oldest and the largest ones, use a subscription-based model, where a user of these databases must agree to a license and is restricted with respect of what he or she may do with the data obtained from the resource. As long as the main vehicle of database dissemination were paper editions or magnetic tape reels that could be used only in computer centers, such situation seemed fairly acceptable. In the epoch of ubiquitous computer access and with the advent of the Internet, however, researchers expressed concerns that certain licensing clauses are overly restrictive. So, the restriction to disseminate derived results was mentioned as an impediment for scientific work (Baldi et al. 2011; Andronico et al. 2011). As a result, several modern databases were created anew, following an open-access dissemination model, and

**Table 1** Overview of largest crystallographic databases, their subject areas, sizes and licensing models

| No. | Database | Records | License | Current URL | Est. | Reference |
|---|---|---|---|---|---|---|
| 1 | PDF | 380,000 | Subscription based | http://www.icdd.com/products/pdf4.htm | 1941 | Faber and Fawcett (2002) |
| 2 | CSD | 800,000 | Subscription based | http://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/ | 1965 | Groom et al. (2016) |
| 3 | PDB | 124,000 | Open access | http://www.rcsb.org/pdb | 1971 | Protein Data Bank (1971) and Berman et al. (2012) |
| 4 | ICSD | 200,000 | Subscription based | https://icsd.fiz-karlsruhe.de/ | 1987 | Belsky et al. (2002) |
| 5 | NDB | 8600 | Open access | http://ndbserver.rutgers.edu/ | 1992 | Berman et al. (1992) and Narayanan et al. (2014) |
| 6 | Pauling file | 290,000 | Subscription based | http://paulingfile.com http://crystdb.nims.go.jp/index_en.html | 1995 | Villars et al. (1998, 2004) |
| 7 | IZA Zeolite database | 176 | Open access | http://www.iza-structure.org/databases/ | 1996 | Baerlocher et al. (2007) |
| 8 | CRYSTMET | 170,000 | Subscription based | http://www.TothCanada.com https://cds.dl.ac.uk/cgi-bin/news/disp?crystmet | 1996 | White et al. (2002) |
| 9 | Bilbao server | | | http://www.cryst.ehu.es | 1997 | Aroyo et al. (2011) |
| 10 | AMCSD | 20,000 | Open access | http://rruff.geo.arizona.edu/AMS/amcsd.php | 2003 | Downs and Hall-Wallace (2003) and Rajan et al. (2006) |

(*continued*)

**Table 1** (continued)

| No. | Database | Records | License | Current URL | Est. | Reference |
|-----|----------|---------|---------|-------------|------|-----------|
| 11 | COD | 367,000 | Public domain | http://www.crystallography.net/cod | 2003 | Gražulis et al. (2009, 2012) |
| 12 | PCOD | 1,000,000 | Public domain | http://www.crystallography.net/pcod | 2003 | Le Bail (2005) |
| 13 | MPOD | 300 | Public domain | http://mpod.cimav.edu.mx | 2010 | Pepponi et al. (2012) |
| 14 | B-IncStrDB (Bilbao Incommensurate Structures Database) | 140 | Open access | http://webbdcrista1.ehu.es/incstrdb/ | 2010 | Aroyo et al. (2006) |
| 15 | TCOD | 2,600 | Public domain | http://www.crystallography.net/tcod | 2013 | Merkys et al. (2017); Chateigner et al. (2015) |
| 16 | RRUFF | 47,000 | Open access | http://rruff.info/ | 2015 | Lafuente et al. (2015) |
| 17 | MAGNDATA (Bilbao Magnetic Structure Database) | 428 | Open access | http://webbdcrista1.ehu.es/magndata/ | 2015 | Perez-Mato et al. (2015) |

in certain cases can be used in situations where licensing requirements are too restricting (Sadowski and Baldi 2013). Among them, the Crystallography Open Database (COD) is currently the largest and the oldest open resource of small molecule crystal structures, providing access to data in mineralogy and chemical crystallography and placing all its collection in public domain.

## 2   A Short History of COD

The COD project started as a community initiative, when crystallographers on the SDPD (Structure Determination by Powder Diffraction) discussed possible modes of crystallographic data dissemination. It was 2003, computers were becoming cheap, Internet connections widely available and free/libre open source software (F/LOSS) ubiquitous. Armel Le Bail raised a question whether it is possible to build an entirely open and free for everyone to use crystallographic database by joining community efforts. Answering that question, Michael Berndt (1964–2003) listed three conditions that were necessary and sufficient for community resource creation and curation: "A small team of engaged scientists with some experience in database and software design to coordinate the project; the authors (i.e., the scientific community = you) who provide the project with database entries /.../; free software (a) for maintaining the database, (b) for data evaluation and calculation of derived data." With this plan in mind, the COD project started and turned out to be a viable alternative to the top-down, heavy-funded database projects. From 2003 to 2007, the COD database master copy was maintained by Armel Le Bail at the Le Mans University in France. In 2007 its collection of 50,000 records was ported to the Institute of Biotechnology in Vilnius, Lithuania, the software development for the COD, and database maintenance was continued. When the Institute of Biotechnology was merged with the Vilnius University in 2011, the COD development continued by the joint team from the Vilnius University Institute of Biotechnology and the Faculty of Mathematics and Informatics.

Despite the several transfers of maintainership, the COD is governed by an international COD Advisory Board (AB), listed on the COD Web site and operating via the mailing list. The COD AB establishes the COD data management policies and sets inclusion criteria for the COD data. In this way, a continuity of database quality is maintained.

During the period of 10 years since 2007, the COD was growing constantly and attained >385,000 records in 2017 (Fig. 1). This was possible with the introduction of the new data deposition Web site (Fig. 2) that allowed both manual and automatic uploads of data to the COD and after development of automated data collection and deposition software that deposits available structures to the COD automatically. This automation in turn is highly facilitated by the introduction of the Crystallographic Interchange Framework (CIF) (Hall et al. 1991; IUCr 2017b). The CIF framework was initially used to facilitate crystallographic paper publication and to reduce typing errors in data by providing automated means of crystallographic data processing (Brown and McMahon 2002). Introduction of electronic data handling
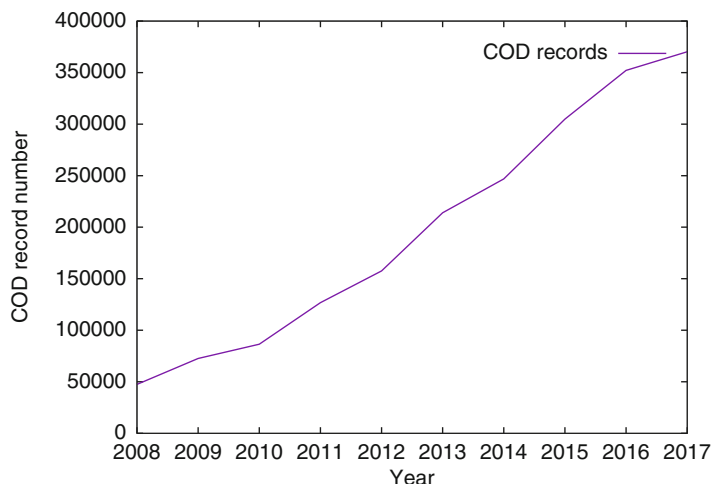
**Fig. 1** Growth of the COD database by year
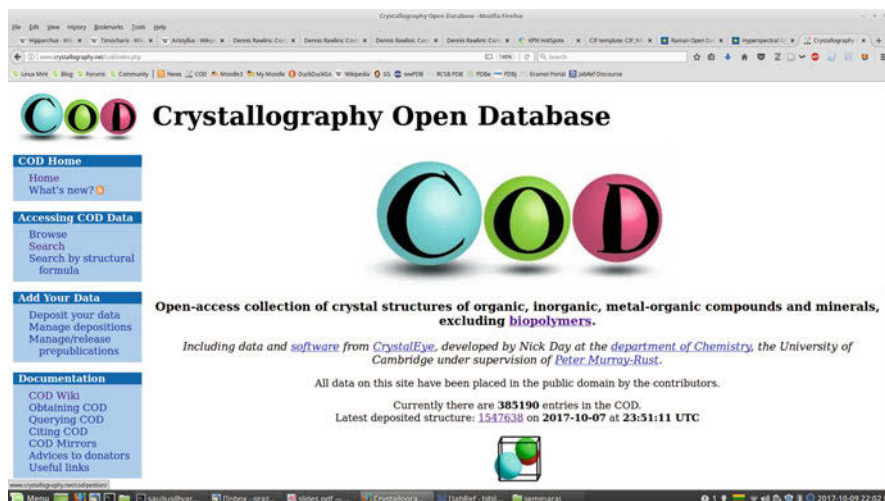


**Fig. 2** The Crystallography Open Database Web site

in the publication process significantly reduced typing errors in data publication, a significant step towards reliable data reuse. Not only that: availability of crystal structure descriptions in a standardized, machine readable form as supplementary material for scientific publications greatly facilitated reuse of that data. As a result the COD data acquisition subsystem can ingest automatically all necessary values and formulate structure description records, using information publicly available with the IUCr publications and from journals of some other publishers that make the necessary information publicly available. The same CIF framework makes

it possible for the COD to present all its data collection in a widely accepted, standardized form, so that researchers can use the same software to process the COD CIFs as for the outputs of structure determination programs or from journal Web pages.

The data collection procedure conducted by the COD is not completely straight-forward, though. Virtually all structures, even though represented in standard CIF format next to the publications, lack essential metadata such as publication bibliography; sometimes computed items such as cell volumes or space group names are missing or presented in a non-standardized form. Such information is automatically inserted by the COD data processing pipeline (Gražulis et al. 2009). Moreover, a non-negligible part of supplementary files, although it does contain necessary data in a form similar to the CIF, does not strictly follow the CIF syntax. Since the number of such cases was too large to be corrected manually, an error-correcting CIF parser was implemented (Merkys et al. 2016). The same procedure is followed when data is deposited by researchers into the COD using the Web deposition interface (Gražulis et al. 2012). In this way the COD ensures that all structure descriptions that enter its collection are syntactically correct, i.e., conform to the syntax defined by the IUCr (2017a).

With this setup, the COD is ready to grow further, to provide open access to crystal structure data for researchers and all interested parties, and to evolve to meet challenges of the new millennium. Computing landscape changes rapidly, with new techniques, languages, formats and protocols coming and going every day, and computer architectures changing fast enough so that any reasonable scientific archive must outlive many generations of computer software and hardware. The basic principles of the COD design and the successful operation of the COD for more than a decade hint that the methods chosen by the COD founders were sound and that the COD will successfully evolve into the future.

## 3    Scope and Contents of the COD

The COD collects machine-readable descriptions of crystal structures for inorganic compounds, minerals, small organic molecules, metal-organic and organometallic compounds. Proteins, nucleic acids and their complexes, glycoproteins and the like are as a rule excluded from the COD, since they are systematically collected in an open-access database, the Protein Data Bank (PDB) (Berman et al. 2012). Most of the "small molecule" structures in the COD are refined using assumption of independent atom parameters (using full-matrix least squares refinement), and a spherical atom model. This makes the COD suitable, for example, to generate restraints on molecular geometries and to refine larger molecules or molecular assemblies (Long et al. 2017a, b). We must note, however, that this assumption does not necessarily hold for all COD entries. For larger entries, or when disorder is present, restraints can be put the by authors on the thermal displacement parameters. For structures solved using powder diffraction techniques, restraints on bond lengths and angles can be also used. Finally, some structures in the COD are solved by

hybrid methods, using powder diffraction to carry out Rietveld refinement and to use DFT to further refine atomic parameters; some structures are reported entirely based on DFT calculations. Obviously, determining bond length and angle parameters from restrained structures would result in circular reasoning, since the same restraints were already used during the structure refinement process. Thus, the user is advised to inspect structure determination parameters and to select those structures that are suitable for his or her work.

## 4    COD Data Semantics and Selection

To facilitate structure selection, the COD maintains a set of flags that describe experimental and refinement techniques used for structure determination. In the COD SQL table, the "`method`" column of the '`data`' table describes the experimental technique which can be "`single crystal`", "`powder diffraction`" or "`theoretical`". If the value of this column is `NULL`, the method is most probably single crystal diffraction. Unfortunately, in many structures the most popular method, "`single crystal`", is not mentioned explicitly, so this assumption is a certain guess; but the structures solved by "`powder diffraction`" or "`theoretical`" methods are usually marked more accurately and are less numerous, so the guess should be reasonably safe. Structures marked as "`theoretical`" are in fact solved by DFT computations without using any structure-specific experimental data. These structures are of course more appropriate to a different database, the TCOD, which is dedicated to theoretical structures, and are in fact also most likely deposited there. They ended up in the COD since they were provided as supplementary material to some papers and were not marked as being theoretical and only later data curation revealed their determination method. Several important theoretical structures, e.g., from the DFT method error estimate studies (Lejaeghere et al. 2014), were deposited to the COD before the TCOD was fully operational but were deemed important enough so that permanent storage in a database for these data records is necessary. Since the COD policy is not to delete any records, so that once assigned COD IDs remain stable, the policy of the COD is to mark its entries with appropriate flags, but not to remove them.

Further the COD database tables contain several fields describing experimental techniques, taken from the IUCr Core CIF dictionary. The "`radiation`", "`radType`" and "`radSymbol`" columns of the "`data`" table are derived directly from the CIF data items `_diffrn_radiation_probe`, `_diffrn_radiation_type` and `_diffrn_radiation_xray_symbol`, respectively. These data items allow distinguishing between structures obtained from X-ray, neutron and electron diffraction data (the "`radiation`" column can have values "`x-ray`", "`neutron`" or "`electron`" for the respective radiation types). Again, like with the "`single crystal`" value, the most popular radiation type, "`x-ray`", is often not marked and thus represented as a `NULL`

value. We can expect that authors are more attentive when they submit a structure made by a less common method, but certain caution is of course appropriate.

When selecting records from the COD, one must keep in mind certain book-keeping data items. Certain structures are deposited to the COD that are deliberately worse than the best possible interpretation; this is usually done in publications to demonstrate that the main interpretation of data offered by authors is correct or indeed the best one. COD policy is to include such structures (so that the paper claims can be easily verified) but to mark them as "suboptimal." In COD CIFs, such structures are marked with `_cod_suboptimal_structure yes` and `_cod_related_optimal_struct` data items, and in the COD `data` table it has a non-NULL value in the "`optimal`" column pointing the related optimal structure. Unless explicit comparison of suboptimal and optimal structures is sought, only structures with NULL "`optimal`" values should be selected.

Another issue is structures that contain known problems. Again, the COD policy is not to remove such structures, once they were included in the COD, but to flag them appropriately. This flag is recorded in the COD database `data` table "`status`" column. Possible values for this column are "`warnings`", "`errors`", and "`retracted`". The "`warnings`" level indicates that the structure might be after all correct but there are strange features, unusual description, or wrong metadata in it. The "`errors`" mark structures that either have been proven wrong by subsequent published observations, authors' corrigenda or contain serious data consistency problems that prevent correct interpretation of the structure. In all cases, `_cod_error_description` gives a human readable description of the problem. Finally, the "`retracted`" in the "`status`" column indicates that the structure was retracted and should not be used under any circumstances. The reasons for retraction may vary, but usually this flag indicates very serious problems up to the outright scientific fraud, as was the case discovered in one IUCr investigation (Harrison et al. 2010); in such cases, the original publications are retracted as well.

The last thing to take care about is the presence of duplicated entries in the COD. Unfortunately, due to less stringent admission procedures in the earliest days of the COD, or due to programming or data encoding errors, sometimes the same structure is deposited more than once to the COD. Once again, when such situation is detected, neither entry is removed from the COD; instead, one entry, usually the most complete one, is declared to be the "main" entry describing this structure, and the others are marked as "duplicates" using the `_cod_duplicate_entry` data item. If the main entry is missing some information that is present in the duplicates, this information is merged into the main entry and committed as a new revision. Duplicate entries are marked by a non-NULL "`duplicateof`" column in the `data` table. Thus, to select only those entries that are not marked as duplicates, one needs to select entries that have "`duplicateof`" column set to NULL.

It must be noted that only technical duplicates are flagged as such in the COD, i.e., only structures that are originating from the same original description and from the same publication. Two structures of the same compound reported in different publications are not considered duplicates and are stored as different COD records.

Even when the same data file is published as supplementary material to two different publications, it is deposited under two different COD identifiers. The rationale here is that a COD record reports an instance of the crystal structure solution reported somewhere, and all such cases must be represented in the database. Further reduction of the multiple records is the responsibility of the COD user, and, indeed, different tasks will require different uniqueness criteria – in some cases these will be based in chemical identity, in other cases on crystal structure identity, and the COD must provide sufficient data for all such queries.

Collecting all above considerations into one SQL query, we can select all non-retracted experimental structures that are not marked as duplicates and have atomic coordinates with a query displayed in Listing 1; the query there reports number of such entries in the current COD SQL database and can be used for further narrowing down the selection based on crystal parameters.

**Listing 1** Number of non-retracted experimental structures with coordinates in the COD that are not marked as duplicates

```bash
#!/bin/bash

mysql -h sql.crystallography.net cod -u cod_reader -t -e \
    'select count(*), current_timestamp() from data
     where duplicateof is null
       and flags like "%has_coordinates%"
       and (status is null or status != "retracted")
       and (method is null or method != "theoretical")'
```

```
+----------+---------------------+
| count(*) | current_timestamp() |
+----------+---------------------+
|   383573 | 2017-12-04 14:07:01 |
+----------+---------------------+
```

## 5  Accessing the COD

### 5.1  Web Access to the COD

The COD offers several methods to access its structure collection. The one that requires least effort to learn is probably to use query forms (see Fig. 3). Multiple parameters can be specified, most of which should be self-explanatory; their exact meaning, however, is the same as in the RESTful query fields and can be looked up in the Table 2.

Results from a Web query are displayed in a separate browser page as a HTML table (Fig. 4); in addition to that, options are provided to download the list of resulting structures as a list of COD identifiers, download URLs or as a CSV format table. For a small number of hits, a ZIP archive of all found CIFs is offered, but for a larger number of structures (typically more than several thousands), this option is

**Fig. 3**  The Web query form of the Crystallography Open Database

**Table 2**  RESTful interface search parameters and their descriptions

| Parameter | Description |
| --- | --- |
| format | The format in which the results will be returned |
| formula | The empirical chemical formula of the crystal. Chemical element symbols in the formula must be ordered according to the Hill notation and separated by a space symbol, i.e., "C8 H10 N4 O2" |
| el1, el2, …, el8 | Chemical element symbols that must appear in the chemical formula |
| nel1, nel2, …, nel4 | Chemical element symbols that must not appear in the chemical formula |
| strictmin, strictmax | The minimum/maximum number of distinct chemical elements that must appear in the chemical formula |
| amin, amax | The minimum/maximum value of the lattice parameter a |
| bmin, bmax | The minimum/maximum value of the lattice parameter b |
| cmin, cmax | The minimum/maximum value of the lattice parameter c |
| minZ, maxZ | The minimum/maximum Z value of the lattice |
| year | The year of publication of the crystal structure |

not available in order to avoid excessive stress on the COD servers, and instead a user is advised to download the COD structures in full and pick the desired CIFs using the COD identifier list resulting from the search.

## 5.2    Using the RESTful Interfaces

The COD offers a RESTful interface that allows one to retrieve information about COD entries based on certain criteria as well as the crystal structure files themselves. The REST (REpresentational State Transfer) is an architectural style of network-based programs that was outlined in the doctoral dissertation of Roy Fielding (2000). The main ideas of this architecture relevant for the COD are to use a client-

**Fig. 4** An example result page from a Crystallography Open Database Web query

server design (the COD server serves multiple clients), to make the COD server stateless as much as possible (thus the same request to the COD server should yield identical results if repeated several times), to use standard connections based on HTTP protocol and stable Web URIs, and to use standard formats (CIF, HTML) to exchange information. An interface based on the ideas of REST, a so-called RESTful interface, has the benefit of not requiring a specialized client program since the queries can be executed by any piece of software capable of resolving URIs including, but not limited to, most Internet browsers.

COD RESTful search query URIs adhere to the HTTP GET query format taking http://www.crystallography.net/cod/result as the basis URI. For example, a query that returns a list of COD IDs associated with structures that contain the Li and O atoms and were published in 2017 would take a form of: http://www.crystallography.net/cod/result?el1=Li&el2=O&year=2017&format=lst

As mentioned above, specialized software is not required, but it can, however, ease the construction of the query strings. An example of the same request rewritten to use the cURL program is given in Listing 2.

**Listing 2** Querying the RESTful interface using cURL

```bash
#!/bin/bash

curl 'http://www.crystallography.net/cod/result' \
-d 'el1=Li' \
-d 'el2=O' \
-d 'year=2017' \
-d 'format=lst'
```

Several more examples of COD RESTful interface queries using cURL are given in listings Listings 3, 4, 5, and 6. Description of the used query parameters is given in Table 2. The full list of supported parameters and formats can be acquired at http://wiki.crystallography.net/RESTful_API/.

**Listing 3**   Count of structures that contain Fe atoms, but no O atoms

```
#!/bin/bash

curl 'http://www.crystallography.net/cod/result' \
-d 'el1=Fe' \
-d 'nel1=O' \
-d 'format=count'
```

**Listing 4**   Information about entries that contain only Fe and N atoms in JSON format

```
#!/bin/bash

curl 'http://www.crystallography.net/cod/result' \
-d 'el1=Fe' \
-d 'el2=N' \
-d 'strictmin=2' \
-d 'strictmax=2' \
-d 'format=json'
```

**Listing 5**   Text file with URLs of entries that have the "C O2" chemical formula

```
#!/bin/bash

curl 'http://www.crystallography.net/cod/result' \
-d 'formula'='C O2'
-d 'format=urls'
```

**Listing 6**   ZIP archive containing CIF files of entries that have cell length between 30 Å and 35 Å and Z number between 3 and 4

```
curl 'http://www.crystallography.net/cod/result' \
-d 'amin=30&amax=35' \
-d 'bmin=30&bmax=35' \
-d 'cmin=30&cmax=35' \
-d 'minZ=3&maxZ=4' \
-d 'format=zip'
```

## 5.3   Querying SQL Database

SQL (Structure Query Language) is arguably the most powerful method of interrogating relational databases and offers more features than the COD Web page or even

than the COD RESTful interface. The Crystallography Open Database offers a read-only access to its data tables so that SQL queries can be carried out by user or by third-party software. Covering SQL language syntax and its use is beyond the scope of this chapter, but numerous textbooks and on-line references of SQL exist, as well as excellent documentation of several F/LOSS implementations of SQL (MySQL is one of them). In this text we provide just a few examples that demonstrate how SQL queries can be used for querying the COD out of the box.

The COD SQL tables are constructed automatically from the COD CIF collection. Tables are updated by the post-commit hooks of the Subversion repository; thus the SQL tables should be always in sync with the CIF collection. In the COD, the dataflow is always from CIFs to the SQL database; thus all changes in tables must be first recorded and versioned in the main repository. Thus, MySQL acts essentially as a fast search cache for the COD, making use of index tables and query optimizer. The COD MySQL `data` table contains also the "svnrevision" column that records Subversion revision from which each row is produced. In addition to that, all COD MySQL tables are dumped nightly in text form and committed to the same Subversion repository as the CIF collection. These archives provide means to reproduce queries that were run some time ago, should this necessity arise for scientific computation reproducibility.

The simplest query counts number of records in the current revision of the COD (Listing 7). A more elaborate form of this query which filters structures that are usually unwanted is provided in the Listing 1. Further examples (Listings 8, 9, and 10) demonstrate how various chemical features can be queried. Specifically, the Listing 9 shows how the COD MySQL server can be queried using regular expressions, an extension of the SQL language. These queries permit selections based on atom chemical types, among other possibilities.

**Listing 7** Number of entries in the COD

```
#!/bin/bash

mysql -h sql.crystallography.net cod -u cod_reader -t -e \
    'select count(*), current_timestamp() from data'
```

```
+----------+---------------------+
| count(*) | current_timestamp() |
+----------+---------------------+
|   387948 | 2017-12-04 14:07:02 |
+----------+---------------------+
```

**Listing 8** DOIs and publication years of structures of cucurbituril

```
#!/bin/bash

mysql -h sql.crystallography.net cod -u cod_reader -t -e \
    'select file, doi, year from data
     where chemname like "%cucurbituril%"'
```

```
+---------+--------------------------+------+
| file    | doi                      | year |
+---------+--------------------------+------+
| 2200062 | 10.1107/S1600536800019498 | 2001 |
| 4320271 | 10.1021/ic015520p        | 2001 |
| 4320272 | 10.1021/ic015520p        | 2001 |
| 4320689 | 10.1021/ic010362n        | 2001 |
| 4320690 | 10.1021/ic010362n        | 2001 |
| 4508668 | 10.1021/cg060062m        | 2006 |
| 4508669 | 10.1021/cg060062m        | 2006 |
+---------+--------------------------+------+
```

**Listing 9**  Number of hydrocarbons

```bash
#!/bin/bash

mysql -h sql.crystallography.net cod -u cod_reader -t -e \
    'select count(*), current_timestamp() from data
    where formula regexp
    "- C[[:digit:]]* H[[:digit:]]* -"'
```

```
+----------+---------------------+
| count(*) | current_timestamp() |
+----------+---------------------+
|     1250 | 2017-12-04 14:07:02 |
+----------+---------------------+
```

**Listing 10**  Five most voluminous MOFs

```bash
#!/bin/bash

mysql -h sql.crystallography.net cod -u cod_reader -t -e \
    'select file, chemname, vol from data
    where chemname like "%MOF%"
    order by vol desc
    limit 5'
```

```
+---------+-------------+--------+
| file    | chemname    | vol    |
+---------+-------------+--------+
| 4111295 | mesoMOF-1   | 122163 |
| 1519417 | Y-ftw-MOF-3 | 111361 |
| 1519416 | Y-ftw-MOF-2 |  64231 |
| 7032763 | MOF-205-NO2 |  27851 |
| 7032762 | MOF-205-NH2 |  27846 |
+---------+-------------+--------+
```

# 6 COD Applications

Even though the COD is not as large as some older crystallographic databases, it has numerous applications due to its open nature. One immediate possibility where the COD excels is teaching. Using the COD one can give students some real-life data search and crystallographic applications, illustrate structures of various compounds, and provide insights into modern chemical research areas (Gražulis et al. 2015). Advantages of the COD are its extremely rapid release cycle (the database is updated daily), permissive license that allows students to download arbitrary parts or even the whole database to their computers, and its availability on the Internet where it can be accessed from or outside the classroom.

Another widely accepted application of the COD is its use for material identification with the help of powder diffraction method and search-match procedure. Largest diffractometer vendors (among them Bruker, PANalytical, Rigaku) have adapted the COD collection for their software and ship it with their equipment, providing regular updates on the COD Web site or on their own pages. Since the COD is an open database, these updates are free of charge for the end users. The COD has currently accumulated enough mineral structures so that it can be used for the SOLSA project (http://solsa-mining.eu), where the database is used, together with other information sources, as a tool for material identification and data dissemination.

In bioinformatics and drug design, the COD is used as a source of open data for restraint libraries (Long et al. 2017a, b). It is also used in DataWarrior (Sander et al. 2015) as one of the sources of chemical information and in the OpenMolecules Web site (http://www.openmolecules.org/). Software testing benefits from large collection of COD data, where different cases need to be examined and data needs to be stored in regression tests. Finally, the COD is used in fundamental research to answer different questions about matter (see, e.g., recent works on MOFs (First and Floudas 2013), hydrogen storage (Breternitz and Gregory 2015), or characterization of 2D materials (Mounet et al. 2018).

# 7 Conclusions

The more than decade-long history of the COD has demonstrated that it is possible to build a lasting, high-quality scientific database using an open-access licensing model. At its current state, the COD is useful for a range of academic and industrial applications. Most importantly, this open database provides everyone with the access to knowledge in its own field of small molecule crystallography. At the same time, there are a lot of obvious improvements that can be done. Clearly the COD needs a more comprehensive data collection. More community organization effort should be done, to involve more people in data correction, collection, and ensuring quality of the COD records. More links with the rest of the Internet data resources should be made, integrating the COD more closely into the Linked Open Data

Cloud. None of these tasks seems to be outside the reach of current possibilities, and so one can expect that in due time, the COD is expanded to include all these features.

# References

Allen FH, Bellard S, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummelink-Peters BG, Kennard O, Motherwell WDS, Rodgers JR, Watson DG (1979) The Cambridge crystallographic data centre: computer-based search, retrieval, analysis and display of information. Acta Crystallogr Sect B Struct Crystallogr Crystal Chem 35(10):2331–2339

Andronico A, Randall A, Benz RW, Baldi P (2011) Data-driven high-throughput prediction of the 3-D structure of small molecules: review and progress. J Chem Inf Model 51:760–776

Aroyo MI, Perez-Mato JM, Capillas C, Kroumova E, Ivantchev S, Madariaga G, Kirov A, Wondratschek H (2006)   Bilbao crystallographic server: I. Databases and crystallographic computing programs. Z Kristallogr – Crystalline Materials 221(1):15–27

Aroyo MI, Perez-Mato JM, Orobengoa D, Tasci E, de la Flor G, Kirov A (2011) Crystallography online: Bilbao crystallographic server. Bulg Chem Commun 43(2):183–197

Baerlocher C, McCusker LB, Olson DH (2007) Atlas of zeolite framework types, 6th edn. Elsevier, Amsterdam/London/New York/Oxford/Paris/Shannon/Tokyo

Baldi P (2011)  Data-driven high-throughput prediction of the 3-D structure of small molecules: review and progress. A response to the letter by the Cambridge crystallographic data centre. J Chem Inf Model 51:3029

Belsky A, Hellenbrandt M, Karen VL, Luksch P (2002) New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. Acta Crystallogr B 58:364–369

Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B (1992) The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J 63:751–759

Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2012) The protein data bank at 40: reflecting on the past to prepare for the future. Structure 20:391–396

Bragg WH (1913) The reflection of x-rays by crystals. (II) Proc R Soc A Math Phys Eng Sci 89(610):246–248

Bragg WH, Bragg WL (1913) The reflection of x-rays by crystals. Proc R Soc Lond A Math Phys Eng Sci 88:428–438

Breternitz J, Gregory D (2015) The search for hydrogen stores on a large scale; a straightforward and automated open database analysis as a first sweep for candidate materials. Crystals 5:617–633

Brown ID, McMahon B (2002) CIF: the computer language of crystallography. Acta Crystallogr B 58:317–324

Chateigner D, Grazulis S, Pérez O, Pepponi G, Lutterotti L (2015)   COD, PCOD, TCOD, MPOD. . . open structure and property databases. http://www.ecole.ensicaen.fr/~chateign/danielc/abstracts/Chateigner_abstract_JNCO2013.pdf accessed 2018-10-03

Clews CJB, Cochran W (1948) The structures of pyrimidines and purines. I. A determination of the structures of 2-amino-4-methyl-6-chloropyrimidine and 2-amino-4,6-dichloropyrimidine by x-ray methods. Acta Crystallogr 1(1):4–11

Downs RT, Hall-Wallace M (2003) The American mineralogist crystal structure database. Am Miner 88:247–250

Faber J, Fawcett T (2002) The powder diffraction file: present and future. Acta Crystallogr B 58(3 Part 1):325–332

Fielding RT (2000) Architectural Styles and the design of network-based software architectures. Ph.D. thesis, University of California, Irvine

First EL, Floudas CA (2013) Mofomics: computational pore characterization of metal-organic frameworks. Microporous Mesoporous Mater 165:32–39

Friedrich W, Knipping P, Laue M (1912a) Interferenzerscheinungen bei Röntgenstrahlen. Eine quantitative Prüfung der Theorie für die Interferenz-Erscheinungen bei Röntgenstrahlen. Bayerische Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Sitzungsberichte, pp 303–322

Friedrich W, Knipping P, Laue M (1912b) Interferenzerscheinungen bei Röntgenstrahlen. Eine quantitative Prüfung der Theorie für die Interferenz-Erscheinungen bei Röntgenstrahlen, II. Bayerische Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Sitzungsberichte, pp 363–373

Gražulis S, Chateigner D, Downs RT, Yokochi AFT, Quirós M, Lutterotti L, Manakova E, Butkus J, Moeck P, Le Bail A (2009) Crystallography open database: an open-access collection of crystal structures. J Appl Crystallogr 42(4):726–729

Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya NR, Moeck P, Downs RT, Le Bail A (2012) Crystallography open database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Res 40(D1):D420–D427

Gražulis S, Sarjeant AA, Moeck P, Stone-Sundberg J, Snyder TJ, Kaminsky W, Oliver AG, Stern CL, Dawe LN, Rychkov DA, Losev EA, Boldyreva EV, Tanski JM, Bernstein J, Rabeh WM, Kantardjieff KA (2015) Crystallographic education in the 21st century. J Appl Crystallogr 48(6):1964–1975

Groom CR, Allen FH (2014) The Cambridge structural database in retrospect and prospect. Angew Chem Int Ed 53:662–671

Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) The Cambridge structural database. Acta Crystallogr B 72(2):171–179

Hall SR, Allen FH, Brown ID (1991) The crystallographic information file (CIF): a new standard archive file for crystallography. Acta Crystallogr A 47(6):655–685

Harrison WTA, Simpson J, Weil M (2010) Editorial. Acta Crystallogr E Struct Rep Online 66(1):e1–e2

Hermann C, Ewald PP (1931) Strukturbericht 1913-1928: Zeitschrift für Kristallographie, Kristallgeometrie, Kristallphysik, Kristallchemie. Akademische Verlagsgesellschaft, Leipzig

IUCr (2017a) A formal grammar for CIF. https://www.iucr.org/resources/cif/spec/version1.1/cifsyntax, accessed 2018-10-03

IUCr (2017b) Crystallographic information framework. https://www.iucr.org/resources/cif, accessed 2018-10-03

IUCr (2017c) Structure reports. https://www.iucr.org/publications/other/structure-reports, accessed 2018-10-03

Kabekkodu SN, Faber J, Fawcett T (2002) New powder diffraction file (pdf-4) in relational database format: advantages and data-mining capabilities. Acta Crystallogr B 58:333–337

Kaduk JA (2002) Use of the inorganic crystal structure database as a problem solving tool. Acta Crystallogr B 58(Pt 3 Pt 1):370–379

Kitaigorodsky AI (1955) Organic Crystal Chemistry (Akad. Nauk SSSR, Moscow) [in Russian]

Lafuente B, Downs RT, Yang H, Stone N (2015) The power of databases: the RRUFF project. In: Highlights in mineralogical crystallography. W. De Gruyter, Berlin, pp 1–30

Le Bail A (2005) Inorganic structure prediction with *GRINSP*. J Appl Crystallogr 38:389–395

Lejaeghere K, Van Speybroeck V, Van Oost G, Cottenier S (2014) Error estimates for solid-state density-functional theory predictions: an overview by means of the ground-state elemental crystals. Crit Rev Solid State Mater Sci 39:1–24

Long F, Nicholls RA, Emsley P, Gražulis S, Merkys A, Vaitkus A, Murshudov GN (2017a) ACEDRG: a stereo-chemical description generator for ligands. Acta Crystallogr D 73(2):112–122

Long F, Nicholls RA, Emsley P, Gražulis S, Merkys A, Vaitkus A, Murshudov GN (2017b) Validation and extraction of stereochemical information from small molecular databases. Acta Crystallogr D 73(2):103–111

Merkys A, Vaitkus A, Butkus J, Okulič-Kazarinas M, Kairys V, Gražulis S (2016) *COD::CIF::Parser*: an error-correcting CIF parser for the Perl language. J Appl Crystallogr 49(1):292–301

Merkys A, Mounet N, Cepellotti A, Marzari N, Gražulis S, Pizzi G (2017) A posteriori metadata from automated provenance tracking: integration of AiiDA and TCOD. J Cheminform 9(1):56

Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, Sohier T, Castelli IE, Cepellotti A, Pizzi G, Marzari N (2018) Novel two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. Nat Nanotechnol 13(3):246–252

Narayanan BC, Westbrook J, Ghosh S, Petrov AI, Sweeney B, Zirbel CL, Leontis NB, Berman HM (2014) The nucleic acid database: new features and capabilities. Nucleic Acids Res 42:D114–D122

Pepponi G, Gražulis S, Chateigner D (2012) MPOD: a material property open database linked to structural information. Nucl Instrum Methods Phys Res Sect B: Beam Interact Mater Atoms 284:10–14. E-MRS 2011 Spring Meeting, Symposium M: X-ray techniques for materials research-from laboratory sources to free electron lasers

Perez-Mato JM, Gallego SV, Tasci ES, Elcoro L, de la Flor G, Aroyo MI (2015) Symmetry-based computational tools for magnetic crystallography. Annu Rev Mater Res 45(1):217–248

Protein Data Bank (1971) Protein data bank. Nat New Biol 233:22–23

Rajan H, Uchida H, Bryan DL, Swaminathan R, Downs RT, Hall-Wallace M (2006) Building the American mineralogist crystal structure database: a recipe for construction of a small internet database. In: Sinha AK (ed) Geoinformatics: data to knowledge, Geological Society of America, Boulder, vol 397, pp 73–80

Röntgen WC (1896) On a new kind of rays. Nature 53:274–276

Sadowski P, Baldi P (2013) Small-molecule 3D structure prediction using open crystallography data. J Chem Inf Model 53:3127–3130

Sander T, Freyss J, von Korff M, Rufener C (2015) DataWarrior: an open-source program for chemistry aware data visualization and analysis. J Chem Inf Model 55(2):460–473

Villars P, Onodera N, Iwata S (1998) The Linus Pauling file (LPF) and its application to materials design. J Alloys Compd 279:1–7

Villars P, Cenzual K, Daams J, Chen Y, Iwata S (2004) Data-driven atomic environment prediction for binaries using the Mendeleev number: part 1. Composition AB. J Alloys Compd 367(1–2):167–175. Proceedings of the VIII international conference on crystal chemistry of intermetallic compounds

White PS, Rodgers JR, Le Page Y (2002) Crystmet: a database of the structures and powder patterns of metals and intermetallics. Acta Crystallogr B 58(Pt 3 Pt 1):343–348

# Quantum Machine Learning in Chemistry and Materials

# 86

Bing Huang, Nadine O. Symonds, and O. Anatole von Lilienfeld

## Contents

**Abstract**

Within the past few years, we have witnessed the rising of quantum machine learning (QML) models which infer electronic properties of molecules and materials, rather than solving approximations to the electronic Schrödinger equation. The increasing availability of large quantum mechanics reference

B. Huang · N. O. Symonds · O. A. von Lilienfeld (✉)
Department of Chemistry, Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), University of Basel, Basel, Switzerland
e-mail: bing.huang@unibas.ch; nsnadinesymonds@gmail.com; anatole.vonlilienfeld@unibas.ch

datasets has enabled these developments. We review the basic theories and key ingredients of popular QML models such as choice of regressor, data of varying trustworthiness, the role of the representation, and the effect of training set selection. Throughout we emphasize the indispensable role of learning curves when it comes to the comparative assessment of different QML models.

# 1    Introduction

Society is becoming increasingly aware of its desperate need for new molecules and materials, be it new antibiotics or efficient energy storage and conversion materials. Unfortunately, chemical compounds reside in, or rather hide among, an unfathomably huge number of possibilities, also known as chemical compound space (CCS). CCS is the set of stable compounds which can be obtained through all combinations of chemical elements and interatomic distances. For medium-sized drug-like molecules, CCS is believed to exceed $10^{60}$ (Kirkpatrick and Ellis 2004). Exploration in CCS and locating the "optimal" compounds are thus an extremely difficult, if not impossible, task. Typically, one needs to constrain the search domain in CCS and obtain certain pertinent properties of compounds within the subspace and then choose the compounds with properties which come closest to some preset criteria as potential candidates for subsequent updating or validation. Of course, one can conduct experiments for each compound. Alternatively, one can also attempt to estimate its properties using modern atomistic simulation tools which, within one approximation or the other, attempt to solve Schrödinger's equation on a modern powerful computer.

The latter approach is practically more favorable and referred as high-throughput (HT) computational screening (Greeley et al. 2006). In spite of its popularity, it is inherently limited by the computational power accessible considering that (1) the number of possible compounds is much larger than what HT typically is capable of dealing with ($\sim 10^3$) and (2) often very time-consuming explicitly electron correlated methods are necessary to reach chemical accuracy (1 kcal/mol for energies), with computational cost often scaling as $O(N^6)$ ($N$ being the number of electrons, a measure of the system size). Computationally more efficient methods generally suffer from rather weak predictive power. They range from force fields and semiempirical molecular orbital methods, density functional theory (DFT) methods to so-called linear scaling methods which assume locality by virtue of fragments or localized orbitals (Kitaura et al. 1999). It remains an outstanding challenge within conventional computational chemistry that efficiency and accuracy apparently cannot coexist.

To tackle this issue, Rupp et al. (2012) introduced a machine learning (ML) Ansatz in 2012, capable of predicting atomization energies of out-of-sample molecules fast and accurately for the first time. By now many subsequent studies showed that ML models enable fast and yet arbitrarily accurate prediction for *any* quantum-mechanical property. This is no "free lunch"; however, the price to pay

consists of the acquisition of a set of pre-calculated training datasets which must be sufficiently representative and dense.

So what is machine learning? It is a field of computer science that gives computers the ability to learn without being explicitly programmed (Samuel 2000). Among the broad categories of ML tasks, we focus on a type called supervised learning with continuous output, which infers a function from labeled training data. Putting it formally, given a set of $N$ training examples of the form $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$ with $x_i$ and $y_i$ being, respectively, the input (the representation) and output (the label) of example $i$, a ML algorithm models the implicit function $f$ which maps input space $X$ to label space $Y$. The trained model can then be applied to predict $y$ for a new input $x$ (belonging to the so-called test set) absent in the training examples. For quantum chemistry problems, the input of QML (also called representation) is usually a vector/matrix/tensor directly obtained from composition and geometry $\{Z_I, \mathbf{R}_I\}$ of the compound, while the label could be *any* electronic property of the system, notably the energy. The function $f$ is implicitly encoded in terms of the nonrelativistic Schrödinger equation (SE) within the Born-Oppenheimer approximation, $\hat{H}\Psi = E\Psi$, whose exact solution is unavailable for all but the smallest and simplest systems. To generate training data, methods with varied degrees of approximation have to be used instead, such as the aforementioned DFT, QMC, etc.

Given a specific pair of $X$ and $Y$, there are multiple strategies to learn the implicit function $f : X \rightarrow Y$. Some of the most popular ones are artificial neural network (ANN, including its various derivatives, such as convolutional neural network) and kernel ridge regression (KRR, or more generally Gaussian process regression).

Based on a recent benchmark paper (Faber et al. 2017), KRR and ANN are competitive in terms of performance. KRR, however, has the great advantage of simplicity in interpretation and ease in training, provided an efficient representation is used. Within this chapter, we therefore focus on KRR or Gaussian processes exclusively (see Sect. 2 for more details.).

Often, each training example is represented by a pair $(x_i, y_i)$. However, multiple $\{y_j\}_i$ can also be used, e.g., when multiple labels are available for the same molecule, possibly resulting from different levels of theory. The latter situation can be very useful for obtaining highly accurate QML models with scarcely available accurate training data and coarse data being easy to obtain. Multi-fidelity methods take care of such cases and will be discussed in Sect. 3.

Once the suitable QML model is selected, be it either in terms of ANN, KRR, or in terms of a multi-fidelity approach, two additional key factors will have a strong impact on the performance: The material representation and the selection procedure of the training set. The representation of any compound should essentially result from a bijective map which uses as input the same information which is also used in the electronic Hamiltonian of the system, i.e., compositional and structural information $\{Z_I, \mathbf{R}_I\}$ as well as electron number. The representation is then typically formatted into a vector which can easily be processed by the computer. Some characteristic representations, introduced in the literature, are described in Sect. 4, where we will see how the performance of QML models can be

enhanced dramatically by accounting for more of the underlying physics. In Sect. 5, further improvements in QML performance are discussed resulting from rational training set selection, rather than from random sampling.

Having introduced the basics of ML, we are motivated to point out two aspects of ML that may not be obvious for better interpretation of how ML works: (1) ML is an inductive approach based on rigorous implementation of inductive reasoning and it does not require *any* a priori knowledge about the aforementioned implicit function $f$ (see Sect. 2), though some insight of what $f$ may look like is invaluable for rational design of representation (see Sect. 4); (2) ML is of interpolative nature, that is, to make reasonable prediction, the new input must fall into the interpolating regime. Furthermore, as more training examples are added to the interpolating regime, the performance of the ML model can be systematically improved for a quantified representation (see Sect. 4).

## 2    Gaussian Process Regression

In this section, we discuss the basic idea of data-driven prediction of labels: the Gaussian process regression (GPR). In the case of a global representation (i.e., the representation of any compound as a single vector, see Sect. 4 for more details), the corresponding QML model takes the same form as in kernel ridge regression (KRR), also termed the global model. GPR is more general than KRR in the sense that GPR is equally applicable to local representations (i.e., the representation of any compound as a 2D array, with each atom in its environment represented by a single vector, see Sect. 4 for more details). Local GPR models can still successfully be applied when it comes to the prediction of extensive properties (e.g., total energy, isotropic polarizability, etc.) which profit from nearsightedness. The locality can be exploited for the generation of scalable GPR-based QML models which can be used to estimate extensive properties of very large systems.

### 2.1    The Global Model

Here we review the Bayesian analysis of the nonlinear regression model (Rasmussen and Williams 2006) with Gaussian noise $\varepsilon$:

$$\mathbf{y} = \phi(\mathbf{x})^\top \mathbf{w} + \varepsilon, \tag{1}$$

where $\mathbf{x} \in \mathbf{X}$ is the representation, $\mathbf{w}$ is a vector of weights, and $\phi(\mathbf{x})$ is the basis function (or kernel) which maps a $D$-dimensional input vector $\mathbf{x}$ into an $N$ dimensional feature space. This is the space into which the input vector is mapped, e.g., for an input vector $\mathbf{x}_1 = (x_{11}, x_{12})$ with $D = 2$, its feature space could be $\phi(\mathbf{x}_1) = (x_{11}^2, x_{11}x_{22}, x_{22}x_{11}, x_{22}^2)$ with $N = 4$. $\mathbf{y}$ is the label, i.e., the observed property of target compounds. We further assume that the noise $\varepsilon$ follows an independent, identically distributed (iid) Gaussian distribution with zero mean

and variance $\lambda$, i.e., $\varepsilon \sim \mathcal{N}(0, \lambda)$, which gives rise to the probability density of the observations given the parameters $\mathbf{w}$, or the likelihood:

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_{i=1}^{n} \mathcal{N}(\phi(\mathbf{x}_i)^\top \mathbf{w}, \lambda I) = \mathcal{N}(\phi(\mathbf{X})^\top \mathbf{w}, \lambda I), \tag{2}$$

where $\phi(X)$ is the aggregation of columns $\phi(\mathbf{x})$ for all cases in the training set. Now we put a zero mean Gaussian prior with covariance matrix $\Sigma_p$ over $\mathbf{w}$ to express our beliefs about the parameters before we look at the observations, i.e., $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$. Together with Bayes' rule

$$p(\mathbf{w}|\mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w})}{p(\mathbf{y}|\mathbf{X})} \tag{3}$$

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w})d\mathbf{w}, \tag{4}$$

distribution of $\mathbf{w}$ can be updated as

$$p(\mathbf{w}|\mathbf{X}, \mathbf{y}) \sim \mathcal{N}(\bar{\mathbf{w}} = \lambda^{-1} A^{-1} \phi(\mathbf{X})\mathbf{y}, A^{-1}) \tag{5}$$

where $A = \lambda^{-1}\phi(\mathbf{X})\phi(\mathbf{X})^\top + \Sigma_p^{-1}$. The updated $\mathbf{w}$ is called the posterior with mean $\bar{\mathbf{w}}$. Thus, similar to Eq. (4), the predictive distribution for $\mathbf{y}_* = f(\mathbf{x}_*)$ is

$$p(\mathbf{y}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \int p(\mathbf{y}_*|\mathbf{x}_*, \mathbf{w})p(\mathbf{w}|X, \mathbf{y})d\mathbf{w}. \tag{6}$$

Substituting Eqs. (2) and (5) into Eq. (6),

$$p(\mathbf{y}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\lambda^{-1}\phi(\mathbf{x}_*)^\top A^{-1}\phi(\mathbf{X})\mathbf{y}, \phi(\mathbf{x}_*)^\top A^{-1}\phi(\mathbf{x}_*)), \tag{7}$$

which can be further simplified to $p(\mathbf{y}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\bar{\mathbf{y}}_*, \bar{\lambda})$ with $\bar{\mathbf{y}}_*$ and $\bar{\lambda}$ being, respectively,

$$\bar{\mathbf{y}}_* = K(\mathbf{x}_*, X)(K(\mathbf{X}, \mathbf{X}) + \lambda I)^{-1}\mathbf{y}, \tag{8}$$

$$\bar{\lambda} = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{X})(K(\mathbf{X}, \mathbf{X}) + \lambda I)^{-1}K(\mathbf{X}, \mathbf{x}_*), \tag{9}$$

where $I$ is the identity matrix and $K(\mathbf{X}, \mathbf{X}) = \phi'(\mathbf{X})^\top \phi'(\mathbf{X})$ $(\phi'(\mathbf{X}) = \Sigma_p^{1/2}\phi(\mathbf{X}))$ is the kernel matrix (also called covariance matrix, abbreviated as Cov). It's not necessary to know $\phi$ explicitly; their existence is sufficient. Given a Gaussian basis function, i.e., $\phi'(x) = \exp(-(x - x_0)^2/(2l^2))$ with $x_0$ and $l$ being some fixed parameters, it can be easily shown that the $(i, j)$-th element of kernel matrix $K$ is

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2}\frac{||\mathbf{x}_i - \mathbf{x}_j||_2^2}{\sigma^2}\right), \tag{10}$$

where $|| \cdot ||_p$ is the $L_p$ norm and $\sigma$ is the kernel width determining the characteristic length scale of the problem. Note that we have avoided the infeasible computation of feature vectors of infinite size by using some kernel function $k$. This is also called the kernel trick. Other kernels can be used just as well, e.g., the Laplacian kernel, $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||_1}{\sigma}\right)$.

Rewriting Eq. 8, we arrive at a more concise expression in matrix form,

$$\mathbf{y}_* = K(\mathbf{X}_*, \mathbf{X})\mathbf{c}, \tag{11}$$

where $\mathbf{c}$ is the regression coefficient vector,

$$\mathbf{c} = (K(\mathbf{X}, \mathbf{X}) + \lambda I)^{-1}\mathbf{y}. \tag{12}$$

Note that this expression can also be obtained by minimizing the cost function $C(\mathbf{w}) = \frac{1}{2}\sum_i (y_i - \mathbf{w}^\top \phi(\mathbf{x}_i))^2 + \frac{\lambda}{2}||\mathbf{w}||_2^2$ with respect to $\mathbf{w}$

Note that $L_2$ regularization is used, together with a regularization parameter $\lambda$ acting as a weight to balance minimizing the sum of squared error (SSE) and limiting the complexity of the model. This eventually leads to a model called kernel ridge regression (KRR) model.

All variants of these global models, however, suffer from the scalability problem for extensive properties of the system such as energy, i.e., the prediction error grows systematically with respect to query system size (predicted estimates will tend toward the mean of the training data while extensive properties grow). This limitation is due to the interpolative nature of global ML models, that is, the predicted query systems and their properties must lie within the domain of training data.

## 2.2 The Local Version

The scalability problem can be overcome by working with local, e.g., atomic, representations. This relies on the idea that one can decompose a global extensive property of the system into local contributions. Among the many ways to partition systems into building blocks, we select the atom-in-molecule (AIM) idea, put forth many years ago by Bader (1990). For the total energy ($E$) of the system, it is usually expressed as a sum over atomic energies ($e$):

$$E = \sum_I e^I = \sum_I \int_{\Omega_I} \langle \Psi | \hat{H} | \Psi \rangle d^3 r \tag{13}$$

where $\Omega_I$ is the atomic basin determined by the zero-flux condition of the electron density,

$$\nabla \rho(\mathbf{r_s}) \cdot \mathbf{n}(\mathbf{r_s}) = 0, \quad \text{for every point } \mathbf{r_s} \text{ on the surface } S(\mathbf{r_s}) \tag{14}$$

where $\mathbf{n}(\mathbf{r_s})$ is the unit vector normal to the surface at $\mathbf{r_s}$. The advantage of using Bader's scheme is that the total energy is exactly recovered and that, at least in principle, it includes all short- and long-ranged bonding, i.e., covalent as well as non-covalent (e.g., van der Waals interaction, Coulomb interaction, etc.). Furthermore, due to nearsightedness of atoms in electronic system (Prodan and Kohn 2005), atoms with similar local chemical environments contribute a similar amount of energy to the total energy. Using the notion of alchemical derivatives, this effect, a.k.a. chemical transferability, has recently been demonstrated numerically (Fias et al. 2017). Thus it is possible to learn effective atomic energies based on a representation of the local atoms. Unfortunately, the explicit calculation of local atoms is computationally involved (the location of the zero-flux plane is challenging for large molecules), making this approach less favorable. Instead, we can also assume that the aforementioned Bayesian model is applicable to atomic energies as well, i.e.,

$$e^I = \phi(\mathbf{x}^I)^\top \mathbf{w} + \varepsilon \tag{15}$$

where $\mathbf{x}^I$ is an atomic representation of atom $I$ in a molecule. By summing up terms on both sides in Eq. 15, we have

$$E = \sum_I \phi(\mathbf{x}^I)^\top \mathbf{w} + \varepsilon. \tag{16}$$

Following Bartók et al. (2010), the covariance of the total energies of two compounds can be expressed as

$$K_{ij} = \text{Cov}(E_i, E_j) = \text{Cov}\left(\sum_I e_i^I, \sum_J e_j^J\right) = \sum_I \sum_J \text{Cov}\left(e_i^I, e_j^J\right)$$
$$= \sum_I \sum_J k\left(\mathbf{x}_i^I, \mathbf{x}_j^J\right) \tag{17}$$

where $I$ and $J$ run over all the respective atomic indices in molecule $i$ and $j$ and where $\mathbf{x}_i^I$ is the representation of atom $I$ in molecule $i$.

By inserting Eq. (17) in Eq. (11), we arrive at the formula for the energy prediction of a molecule $*$ out-of-sample:

$$E_* = \sum_i c_i \sum_{I \in i} \sum_{J \in *} k\left(\mathbf{x}_i^I, \mathbf{x}_*^J\right) \tag{18}$$

where $c_i = \sum_j ([K + \lambda I]^{-1})_{ij} E_j$. This equation can be rearranged:

$$E_* = \sum_{J \in *} \sum_i c_i \sum_{I \in i} k\left(\mathbf{x}_i^I, \mathbf{x}_*^J\right) = \sum_{J \in *} e_*^J, \tag{19}$$

where the atomic contribution of atom $J$ to the total energy can be decomposed into a linear combination of contributions from each training compound $i$, weighted by its regression coefficient,

$$e_*^J = \sum_i c_i \tilde{e}_{*i}^J. \tag{20}$$

The "basis function" $\tilde{e}_{*i}^J$ in this expansion simply consists of the sum over kernel similarities between atom $J$ and atoms $I \in i$, where the contribution of atom $I$ grows with its similarity to atom $J$:

$$\tilde{e}_{*i}^J = \sum_I k\left(\mathbf{x}_i^I, \mathbf{x}_*^J\right). \tag{21}$$

We note in passing that the value of the covariance matrix element (i.e., Eq. (17)) increases when the size of either system $i$ or $j$ grows, indicating that the scalability issue can be effectively resolved.

## 2.3 Hyper-Parameters

Within the framework of GPR or KRR, there are two sets of parameters: (1) parameters that are determined via training, i.e., the coefficients $\mathbf{c}$ (see Eq. (12)), whose number grows with the training data, and (2) hyper-parameter whose value is set before the learning process begins, i.e., the kernel width $\sigma$ in Eq. (11) and $\lambda$ in Eq. (2).

As defined in Sect. 2.1, $\lambda$ measures the level of noise in the training data in GPR. Thus, if the training data is noise free, $\lambda$ can be safely set to zero or a value extremely close to zero (e.g., $1 \times 10^{-10}$) to reach optimal performance. This is generally true for datasets obtained by typical quantum chemical calculations, and the resulting training error is (almost) zero. Whenever there is noise in the data (e.g., from experimental measurements), the best $\lambda$ corresponds to some finite value depending on the noise level. The same holds for the training error. In terms of KRR, $\lambda$ seems to have a completely different meaning at first glance: the regularization parameter determining the complexity of the model. In essence, they amount to the same, i.e., a minute or zero $\lambda$ corresponds to the perfectly interpolating model which connects every single point in the training data, thus representing the most faithful model for the specific problem at hand. One potential risk is poor generalization to new input data (test data), as there could be "overfitting" scenarios for training sets. A finite $\lambda$ assumes some noise in the training data, and the model can only account for this in an averaged way; thus the model complexity is simplified to some extend by

lowering the magnitude of parameters $\mathbf{w}$ so as to minimize the cost function $C(\mathbf{w})$. Meanwhile, some finite training error is introduced. To recap, the balance between SSE and regularization is vital and reflected by a proper choice of $\lambda$.

Unlike $\lambda$, the optimal value of $\sigma$ ($\sigma_{\text{opt}}$) is more dataset specific. Roughly speaking, it is a measure of the diversity of the dataset and controls the similarity (covariance matrix element) of two systems. Typically $\sigma_{\text{opt}}$ gets larger when the training data expands into a larger domain. The meaning of $\sigma$ can be elaborated by considering two extremes: (1) when $\sigma$ approaches zero, the training data will be reproduced exactly, i.e., $c_i = y_i$, with high error for test data, i.e., with deviation to mean, and (2) when $\sigma$ is infinity, all kernel matrix elements will tend toward one, i.e., a singular matrix, resulting in large errors in both training and test. Thus, the optimal $\sigma$ can be interpreted as a coordinate scaling factor to render the kernel matrix well-conditioned. For example, Ramakrishnan and von Lilienfeld (2015) selected the lower bound of the kernel matrix elements to be 0.5. For a Gaussian kernel, this implies that $K_{\text{min}} = \exp\left(-D_{\text{max}}^2/2\sigma_{\text{opt}}^2\right) \approx 0.5$, or $\sigma_{\text{opt}} \approx D_{\text{max}}/\sqrt{2\ln 2}$, where $D_{\text{max}}$ is the largest distance matrix element of the training data. Following the same reasoning, $\sigma_{\text{opt}}$ can be set to $D_{\text{max}}/\ln 2$ for a Laplacian kernel.

The above heuristics are very helpful to quickly identify reasonable initial guesses for hyper-parameters for a new dataset. Subsequently, the optimal values of the hyper-parameters should be fine-tuned through $k$-fold cross-validation (CV). The idea is to first split the training set into $k$ smaller sets, and (1) for each of the $k$ subsets, a model is trained using the remaining $k-1$ subsets as training data; the resulting model is tested on the remaining part of the data to calculate the predictive error); this step yields $k$ predictions, one for each fold. (2) The overall error reported by $k$-fold cross-validation is then the average of the above $k$ values. The optimal parameters will correspond to the ones minimizing the overall error. This approach can become computationally demanding when $k$ and the training set size are large. But it is of major advantage in problem such as inverse inference where the number of samples is very small, and its systematic applications minimize the likelihood of statistical artifacts.

## 2.4   Learning Curves

To assess the predictive performance of a ML model, we need to know not only the prediction error ($\varepsilon$, which can be characterized by the mean absolute error (MAE) or root mean squared error (RMSE) of prediction) for a specific training set but also predictive errors for varied sizes of training sets. Therefore, we can monitor how much progress we have achieved after some incremental changes to the training set size ($N$) so as to extrapolate to see how much more training data is needed to reach a desirable accuracy. The plot of $\varepsilon$ versus $N$ relationship is called the learning curve (LC), and examples are shown in Fig. 1 (note that only test error, i.e., MAE for the prediction of new data in test set, is shown; training errors are always zero or minute for noise-free training data).

**Fig. 1** Three representative learning curves with distinguished relative performance
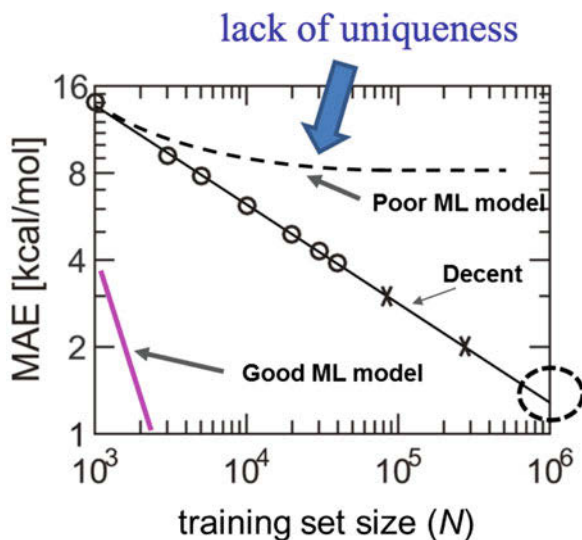
Multiple factors control the shape of learning curve, one of which is the choice of representation. If the representation cannot uniquely encode the molecule, i.e., there may exist cases that two different molecules share the same input vector $\mathbf{x}_i$ but with different molecular properties, then it causes ambiguity to the ML algorithm (see more details in Sect. 4.1) and may consequently lead to no learning at all, as illustrated by the dashed curve in Fig. 1, with distinguishable flattening out behavior at larger training set sizes, resulting in poor ML performance.

In the case of a unique representation, according to Fasshauer and McCourt (2016), it can be proved that for kernel-based approximation, when the training set size $N$ is sufficiently large, the predictive error is proportional to the so-called "fill distance" or *mesh norm* $h_{\mathbf{X}}$, defined as

$$h_{\mathbf{X}} = \sup_{\mathbf{x} \in \Omega} \min_{\mathbf{x}_j \in \mathbf{X}} ||\mathbf{x} - \mathbf{x}_j||_2 \tag{22}$$

where "sup" stands for the supremum (or the least upper bound) of a subset, $\mathbf{x}$ is again the representation of any training instance as an element of the training set $\mathbf{X}$, and $\Omega$ represents the domain of studied systems (i.e., potential energy surface domain for chemistry problems). Clearly from the definition, fill distance describes the geometric relation of the set $\mathbf{X}$ to the domain $\Omega$ and quantifies how densely $\mathbf{X}$ covers $\Omega$. Furthermore, fill distance intrinsically contains a dimension dependence $d$, that is, $h_{\mathbf{X}}$ scales roughly as $N^{-1/d}$ if $\mathbf{x}$ are uniform or random grid points in a $d$ dimensional space.

Apart from the exponent, there should also be a prefactor; thus the leading term of the overall predictive error can be described as $b * N^{-a/d}$, where $a$ in the exponent is a constant. Therefore, to visualize the error vs. $N$, a log-log scale is the most convenient for which the learning curve can be represented by a linear

relationship: $\log(\varepsilon) \approx \log(b) - \frac{a}{d} \log(N)$; thus $a/d$ quantifies the rate of learning, while the prefactor $\log(b)$ is the vertical offset of the learning curve. Through a series of numerical calculations of learning a 1D Gaussian function as well as ground state properties of molecules with steadily improving physics encoded in the representation, it has been found (Huang and von Lilienfeld 2016) that the offset $\log(b)$ is a measure of target property similarity, which is defined as the deviation of proposed model (corresponding to the representation used) from the true model (Huang and von Lilienfeld 2016). While, in general, we do not know the true function (machine learning would be meaningless if we did), we often do have considerable knowledge about relative target similarity of different representations.

Applying the findings above to chemistry problems, we can thus obtain some insight in how learning curves will behave. Several observations can be explained: First, the learning rate would be almost a constant or changes very little when different unique representations are used, as the rate depends primarily on the domain spanned by molecules considered in the potential energy surface. Secondly, for a series of isomers, it is much easier to learn their properties in their relaxed equilibrium state than in a distorted geometry.

The limitation that the learning rate will not change much for random sampling with unique representations seems to be a big obstacle toward more efficient ML predictions, meaning that developing better representation (to lower the offset) can become very difficult even if substantial effort has been invested. However, is it possible to break this curse, reaching an improved learning curve as illustrated by the pink line in Fig. 1? We believe that this should be possible. Note how the linear (log-log) learning curve is obtained for *statistical* models. This implies that there must be "redundancy" in the training data; and if we were able to remove those redundancies a priori, we might very be able to boost the performance and observe superior LCs, such as the pink line in Fig. 1 with large learning rates. In such a case, statistics is unlikely to hold, and the LC may be just a monotonically decreasing function, possibly also just a damped oscillator, rather than a line. Strategies for rational sampling will be elaborated in detail in Sect. 5.

## 3   Multilevel Learning

By default, we assume for each $\mathbf{x}_i \in \mathbf{X}$, there exists one corresponding $y_i \in Y$ in the training examples. It makes perfect sense if $Y$ is easy to compute, i.e., in the circumstance that a relatively low accuracy of $Y$ suffices (e.g., PBE with a medium-sized basis set). It is also possible that a highly accurate reference data is required (e.g., CCSD(T) calculations with a large basis set) so as to achieve highly reliable predictions. Unfortunately, we can only afford few highly accurate $\mathbf{x}$ and $y$'s for training considering the great computational burden. In this situation, one can take great advantage of the $y$'s with lower levels of accuracy which are much easier to obtain. Models which shine in this kind of scenario are called multi-fidelity, where reference data based on a high (low) level of theory is said to have high (low) fidelity.

The nature of this approach is to explore and exploit the inherent correlation among datasets with different fidelities. Here we employ Gaussian process as introduced in Sect. 2 to explain the main concepts and mathematical structure of multilevel learning.

### 3.1 Multi-fidelity

For the sake of clarity and simplicity, we focus only on two levels of fidelity, the mathematical formulation stated below can be easily generalized to more fidelities. We consider two datasets with different level of fidelity: $\{\mathbf{X}, \mathbf{y}^{(1)}\}$ (in which the pairs of data are $(\mathbf{x}_1, y_1^{(1)})$, $(\mathbf{x}_2, y_2^{(1)})$, $\dots$) and $\{\mathbf{X}, \mathbf{y}^{(2)}\}$, where $\mathbf{y}^{(2)}$ has a higher level of fidelity. The number of data points in the two sets is, respectively, $N_1$ and $N_2$ and $N_1 > N_2$, reflecting the fact that high-fidelity data are scarce. We consider the following autoregressive model proposed by Kennedy and O'Hagan (2000):

$$\mathbf{y}^{(2)} = \rho \mathbf{y}^{(1)} + \delta^{(2)} \tag{23}$$

where $\mathbf{y}^{(1)}$ and $\delta^{(2)}$ are two independent Gaussian processes, i.e.,

$$\mathbf{y}^{(1)} \sim \mathcal{N}(0, K_1(\mathbf{X}, \mathbf{X})) = \mathcal{N}(0, \text{Cov}(\mathbf{y}^{(1)}, \mathbf{y}^{(1)})) = \mathcal{N}(0, K_1) \tag{24}$$

$$\delta^{(2)} \sim \mathcal{N}(0, K_2(\mathbf{X}, \mathbf{X}) = \mathcal{N}(0, \text{Cov}(\delta^{(2)}, \delta^{(2)})) = \mathcal{N}(0, K_2). \tag{25}$$

That $\mathbf{y}^{(1)}$ and $\delta^{(2)}$ are independent (notated as $\mathbf{y}^{(1)} \perp \delta^{(2)}$) indicates that the mean of $\mathbf{y}^{(1)}\delta^{(2)}$ satisfies $\mathbf{E}[\mathbf{y}^{(1)}\delta^{(2)}] = \mathbf{E}[\mathbf{y}^{(1)}]\mathbf{E}[\delta^{(2)}]$, and thus the covariance between $\mathbf{y}^{(1)}$ and $\delta^{(2)}$ is zero, i.e., $\text{Cov}(\mathbf{y}^{(1)}, \delta^{(2)}) = \mathbf{E}[\mathbf{y}^{(1)}\delta^{(2)}] - \mathbf{E}[\mathbf{y}^{(1)}]\mathbf{E}[\delta^{(2)}] = 0$. Therefore, $\mathbf{y}^{(2)}$ is also a Gaussian process with mean 0 and covariance:

$$\text{Cov}(\mathbf{y}^{(2)}, \mathbf{y}^{(2)}) = K_{22} = \text{Cov}(\rho \mathbf{y}^{(1)} + \delta^{(2)}, \rho \mathbf{y}^{(1)} + \delta^{(2)}) \tag{26}$$

$$= \rho^2 \text{Cov}(\mathbf{y}^{(1)}, \mathbf{y}^{(1)}) + \text{Cov}(\delta^{(2)}, \delta^{(2)}) = \rho^2 K_1 + K_2 \tag{27}$$

that is, $\mathbf{y}^{(2)} \sim \mathcal{N}(0, \rho^2 K_1 + K_2)$.

The most important term in multi-fidelity theory is the covariance between $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$, which represents the inherent correlation between datasets with different levels of fidelity and is derived as $\text{Cov}(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}) = K_{12} = \rho \text{Cov}(\mathbf{X}, \mathbf{X}) = \rho K_1$ due to the same independence restriction. Now the multi-fidelity structure can be written in the following compact form of a multivariate Gaussian process:

$$\begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}\right), \tag{28}$$

where $K_{11} = K_1$, $K_{22} \neq K_2$, $K_{12} = K_{21}$ due to symmetry. The importance of $\rho$ is quite evident from the term $K_{12}$; specifically, when $\rho = 0$, the high-fidelity and

low-fidelity models are completely decoupled, and there will be no improvements of the prediction at all by combining the two models.

The next step is to make prediction of $\mathbf{y}_*^{(2)}$ given the corresponding input vector $\mathbf{x}_*$, two levels of training data $\{\mathbf{X}, \mathbf{y}^{(1)}\}$ and $\{\mathbf{X}, \mathbf{y}^{(2)}\}$. To this end, we first write down the following joint density:

$$
\begin{pmatrix} \mathbf{y}_*^{(2)} \\ \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix} \sim \mathcal{N}\left( 0, \begin{pmatrix} K_{**} & K_{*1} & K_{*2} \\ K_{1*} & K_{11} & K_{22} \\ K_{2*} & K_{21} & K_{22} \end{pmatrix} \right),
\tag{29}
$$

where $K_{**} = \rho^2 K_1^* + K_2^*$, $K_{*1} = \rho K_1^*$ with $K_1^* = K_1(\mathbf{X}_*, \mathbf{X}_*) = \mathrm{Cov}(\mathbf{y}_*^{(1)}, \mathbf{y}_*^{(1)})$ and $K_2^* = K_2(\mathbf{X}_*, \mathbf{X}_*) = \mathrm{Cov}(\delta_*^{(2)}, \delta_*^{(2)})$; then following similar procedures as in Sect. 2.1, the final predictive distribution of $\mathbf{y}_*^{(2)}|\mathbf{X}_*, \mathbf{X}, \mathbf{y}^{(1)}, \mathbf{y}^{(2)}$ is again a Gaussian $\mathcal{N}(\bar{\mathbf{y}}_*^{(2)}, \mathrm{Var})$, where

$$
\bar{\mathbf{y}}_*^{(2)} = K_* K^{-1} Y, \ \ \mathrm{Var} = K_* K_*^\top - K_* K^{-1} K_*^\top,
\tag{30}
$$

$$
Y = \begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix}, \ \ K_* = \begin{pmatrix} K_{*1} & K_{*2} \end{pmatrix}, \ \ K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}.
\tag{31}
$$

We note in passing that since there are two correlation functions $K_1$ and $K_2$, two sets of hyper-parameters regarding the kernel width and an extra scaling parameter $\rho$ have to be optimized following the similar approach as explained in Sect. 2.4. This algorithm has already successfully been applied to the prediction of band gaps of elpasolite compounds with high accuracy (Pilania et al. 2017). But it can be naturally extended to other properties. So far, not much work has been done using this algorithm; its potential to tackle complicated chemical problems has yet to be unraveled by future work.

## 3.2   Δ-Machine Learning

A naive version of multi-fidelity learning is the so-called $\Delta$-machine learning model. Its performance is useful for the prediction of various molecular properties (Ramakrishnan et al. 2015a). In this model, $N_1$ is equal to $N_2$, the low- and high-fidelity models are, respectively, called baseline and target. The baseline property $(y^{(b)})$ is associated with baseline geometry as encoded in its representation $(\mathbf{x}^{(b)})$, and target property $y^{(t)}$ is associated with target geometry $\mathbf{x}^{(t)}$, respectively. The workhorse of this model is

$$
y_*^{(t)} = y_*^{(b)} + \sum_{i=1}^{N} c_i k(\mathbf{x}_*^{(b)}, \mathbf{x}_i^{(b)})
\tag{32}
$$

Note that we did not use the target geometry at all for the reason that (1) it is expensive to calculate and (2) it is not necessary for the test molecules.

The $\Delta$-ML model has been shown to be capable of yielding highly accurate results for energies if a proper baseline model is used. Other properties can also be predicted with much higher precision compared to traditional single fidelity model (Ramakrishnan et al. 2015a). What is more, this approach can save substantial computational time. However, the $\Delta$-machine learning model is not fully consistent with the multi-fidelity model. The closest scenario is that we set $K_1 = K_2$ when evaluating kernel functions in Eq. (31), but this will result in something still quite different. There are further issues one would like to resolve, including that (i) the coupling between different fidelities is not clear and that the correlation is rather naively accounted for through the $\Delta$ of the properties from two levels, assuming a smooth transition from one property surface (e.g., potential energy surface) from one level of theory to another. This is questionable and may fail terribly in some cases; (ii) it requires the same amount of data for both levels, which can be circumvented by building recursive versions.

## 4 Representation

The problem of how to represent a molecule or material has been a topic dating back to many decades ago, and the wealth of information (and opinions) about this subject is well manifested by the collection of descriptors compiled in Todeschini and Consonni's Handbook of molecular descriptors (Todeschini and Consonni 2008). According to these authors, the molecular descriptor is defined as "the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment." While the majority of these descriptors are graph-based and used for quantitative structure and activity relationships (QSAR) applications (typically producing rather rough correlation between properties and descriptor), our focus is on QML models, i.e., physics-based, systematic, and universal predictions of well-defined quantum-mechanical observables, such as the energy von Lilienfeld (2018). Thus, to better distinguish the methods reviewed here-within from QSAR, we prefer to use the term "representation" rather than "molecular descriptor." Quantum mechanics offers a very specific recipe in this regard: A chemical system is defined by its Hamiltonian which is obtained from elemental composition, geometry, and electron number exclusively. As such, it is straightforward to define the necessary ingredients for a representation: It should be some vector (or fingerprint) which encodes the compositional and structural information of a given neutral compound.

### 4.1 The Essentials of a Good Representation

There are countless ways to encode a compound into a vector, but what representation can be regarded as "good"? Practically, a good representation should lead to

a decent learning curve, i.e., error steadily decreases as a function of training set size. Conceptually, it should fulfill several criteria, including primarily uniqueness (non-ambiguity), compactness, and being size-extensive (von Lilienfeld et al. 2015).

Uniqueness (or being nonambiguous) is indispensable for ML models. We consider a representation to be unique if there is no pair of molecules that produces the same representation. Lack of uniqueness would result in serious consequences, such as ceasing to learn at an early stage or no learning at all from the very beginning. The underlying origin is not hard to comprehend. Consider two representation vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ for two compounds associated with their respective properties $y_1$ and $y_2$. Now suppose $\mathbf{x}_1 = \mathbf{x}_2$ while $y_1 \neq y_2$ (no degeneracy is assumed). One extreme case is that only these two points are used when training the ML model; obviously we will encounter a singular kernel matrix with all elements being 1; huge prediction errors will result, and basically there is no learning. Even if molecules like these are not chosen for training, it should be clear that such a representation introduces a severe and systematic bias. Furthermore, when trying to predict $y_1$ and $y_2$ after training, the estimate will be the same as the input to the machine is the same. The resulting test error is therefore directly proportional to their property difference.

The compactness requires atom index permutation and rotational and translational invariance, i.e., all redundant degrees of freedom of the system should be removed as much as possible while retaining the uniqueness. This can lead to a more robust representation, meaning (1) the size of training set needed may be significantly reduced and (2) the dimension of the representation vector (thus the size) is minimized, a virtue which becomes important when the necessary training set size becomes large.

Being size-extensive is crucial for prediction of extensive properties, among which the most important is the energy. This leads to the so-called atomic representation or local representation of an atom in a compound. The local unit atom can also consist of bonds, functional groups, or even larger fragments of the compound. As pointed out in Sect. 2.2, this type of representation is the crucial stepping stone for building scalable machine learning models. Even intensive properties such as HOMO-LUMO gap, which typically do not scale with system size, can be modeled within the framework of atomic representations, as illustrated using the Re-Match metric (De et al. 2016). For specific problems, such as force predictions, an analytic form of representation is desirable for analysis and rapid evaluation and for subsequent differentiation (with respect to nuclear charges and coordinates) so as to account for response properties.

## 4.2   Rational Design

It is not obvious how to obtain an optimal representation. In order to obtain a good representation, one has to gain intensive knowledge about the system and structure-property relationship. The use of simplified approximations to solutions of Schrödinger's equation is particularly powerful. The most approximative, yet atomistic, models of SE are universal force fields (FF) which typically reproduce the essential physics for certain system classes, such as bioorganic molecules,

reasonably well. Namely, the atom-pairwise two-body interactions in force fields typically decay as $1/R^n$ ($R$ being the internuclear distance and $n$ being some integer), while 3- and 4-body parts behave as periodic functions of angle and dihedral angle (modern force field approaches also include 2- to $(n-1)$-body interaction in $n$-body interactions). FFs are essentially a special case of the more general many-body expansion (MBE) in interatomic contributions, i.e., an extensive property of the system (e.g., total energy) is expanded in a series of many-body terms, namely, 1-, 2- and 3-body terms, $\cdots$, i.e.,

$$E(\{\mathbf{R}_I\}) = \sum_I^{[Z]} E^{(1)}(R_I) + \sum_{J>I}^{[Z]} E^{(2)}(R_{IJ}) + \sum_{K>J>I}^{[Z]} E^{(3)}(R_{IJ}, R_{IK}, \theta_{IJK}) + \cdots$$

(33)

where $E^{(n)}$ is the $n$-body interaction energy, $R_{IJ}$ is the interatomic distance between atom $I$ and $J$, and $\theta_{IJK}$ is the angle spanned by two vectors $\mathbf{R}_{IJ}$ and $\mathbf{R}_{IK}$. Other important properties can also be expressed in a similar fashion.

By utilizing the basic variables in MBE, including distance, angles, and dihedral angles in their correct physics-based functional form (e.g., the aforementioned $1/R^n$ dependence of 2-body interaction strength), one can already build some highly efficient representations such as BAML and SLATM (*vide infra*). This recipe relies heavily on preconceived knowledge about the physical nature of the problem.

## 4.3 Numerical Optimization

It is possible that for some systems and properties, one does not know which features are of primary importance. And it is not an option to try all features one-by-one considering that there are so many possibilities. In such a situation, the least absolute shrinkage and selection operator (LASSO) can offer suitable relief. LASSO is basically a regression analysis method. Consider a simple linear model: the property of a system is a linear functions of its features, i.e., $\mathbf{y} = \mathbf{X}\mathbf{c}$, where $\mathbf{X}$ is a matrix with each of the $N$ rows being the descriptor vector $\mathbf{x}_i$ of length $D$ for each training data points, $\mathbf{c}$ is the $D$-dimensional vector of coefficients, and $\mathbf{y}$ is the vector of training properties with the $i$-th property being $y_i$. Our task is to find the tuple of features that yields the smallest sum of squared error: $||\mathbf{y} - \mathbf{X}\mathbf{c}||_2^2$. Within LASSO, it is equivalent to a convex optimization problem, i.e.,

$$\underset{\mathbf{c} \in \mathbb{R}^D}{\operatorname{argmin}} ||\mathbf{y} - \mathbf{X}\mathbf{c}||_2^2 + \lambda ||\mathbf{c}||_1$$

(34)

where the use of $L_1$ norm of regularization term is pivotal, i.e., smaller $L_1$ norm can be obtained when larger $\lambda$ is used, thereby purging features of lesser importance. This approach has been exemplified for the prediction of relative crystal phase stabilities (rock-salt vs. zinc-blende) in a series of binary solids (Ghiringhelli et al. 2015). Unfortunately, this approach is limited in that it works best for rather low-

dimensional problems. Already for typical organic molecules, the problem becomes rapidly intractable due to coupling of different degrees of freedom. Under such circumstances, it appears to be more effective to adhere to the aforementioned rational design based heuristics, as manifested by the fact that almost all of the ad hoc representations in the literature are based on manual encoding.

## 4.4   An Overview of Selected Representations

Over the years, numerous molecular representations have been developed by several research groups working on QML. It's not our focus to enumerate all of them but to list and categorize the popular ones. Two categories are proposed; one is based on many-body expansions in vectorial or tensorial form, such as Coulomb matrix (CM), Bag of Bonds (BoB), Bond, Angle-based Machine Learning (BAML), Spectrum of London and Axilrod-Teller-Muto potential (SLATM), and the alchemical and structural radial distribution-based representation introduced by Faber, Christensen, Huang, and von Lilienfeld (FCHL). The other category is an electron density model-based representation called smooth overlap of atomic positions (SOAP).

### 4.4.1   Many-Body Potential-Based Representation

The Coulomb matrix (CM) representation was first proposed in the seminal paper by Rupp et al. (2012). It is a square atom-by-atom matrix with off diagonal elements corresponding to the nuclear Coulomb repulsion between atoms, i.e., $CM_{IJ} = Z_I Z_J / R_{IJ}$ for atom index $I \neq J$. Diagonal elements approximate the electronic potential energy of the free atom, which is encoded as $-0.5 Z_I^{2.4}$. To enforce invariance of atom indexing, one can sort the atom numbering such that the sum of $L_2$ and $L_1$ norm of each row of the Coulomb matrix descends monotonically in magnitude. Symmetrical atoms will result in the same magnitude. A slight improvement over the original CM can be achieved by varying the power low of $R_{IJ}$ (Huang and von Lilienfeld 2016). Best performance is found for an exponent of 6, reminiscent of the leading order term in the dissociative tail of London dispersion interactions. Thus, the resulting representation is also known as London matrix (LM). The superiority of LM is attributed to a more realistic trade-off between the description of more localized covalent bonding and long-range intramolecular non-covalent interactions (Huang and von Lilienfeld 2016).

In spite of the great virtue of uniqueness encoded in CM, it generally suffers from a high offset of learning curve (see Fig. 3). In contrast, the bag-of-bond (BoB) representation (Hansen et al. 2015), a bagged (vectorial) stripped down version of the CM, turns out to result in learning curves with lower offset than CM (see Fig. 3). The BoB representation is a 1-D array, constructed as the concatenation of a series of bags (1-D arrays as well); each corresponds to a specific type of atomic pair, e.g., all C-O pairs (covalently and non-covalently bonded) in the molecule are grouped into the bag labeled as CO, similarly for all other combinations of elemental pairs. Each bag thus includes a set of nuclear Coulomb repulsion values. Each bag is then sorted in descending order. In cases that the same type of bag for two molecules has not the
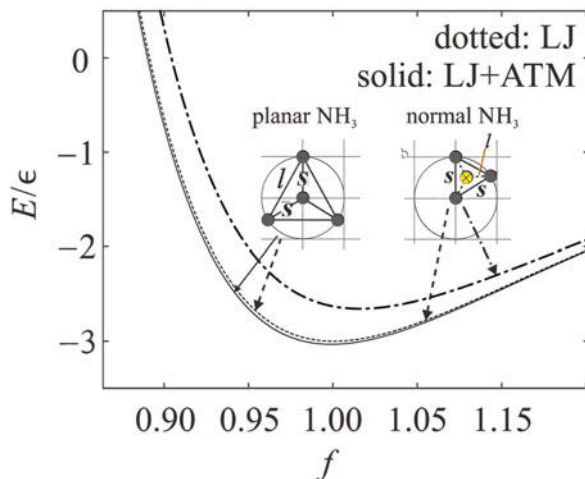
**Fig. 2** Two-body interaction is not enough to capture the physics of a pair of homometric molecules. In the figure, the energy of the two molecules is approximated as summation of LJ potentials with (dashed lines) or without three-body ATM potentials (solid line) and plotted as a function of $f$, the scaling factor of all coordinates of the two molecules. LJ and ATM stand for Lennard-Jones and Axilrod-Teller-Muto vdW potential, respectively. The letters $s$ and $l$ label the two existing different bond lengths, standing for "short" and "long." The atom represented by a yellow-filled circle with cross means out of plane

same size, the smaller bag is padded with zeros. Through bagging the performance is improved in comparison to the CM matrix. But inevitably, crucial higher-order information, such as the angular part, is missing. Due to its exclusive reliance on sorted two-body terms, BoB is not a unique representation, as also manifested by the deterioration of its slope in the learning curve for large training set sizes (see Fig. 3). This loss of information can also be illustrated for a pair of homometric molecules (same atom types, same set of interatomic distances) as displayed in Fig. 2. If we make a plot of the potential energy (approximated as a sum of Lennard-Jones potentials) curve of both planar and tetrahedral molecules as a function of the scaling factor $f$ of all coordinates, we will end up with the same curve due to a spurious degeneracy imposed by lack of uniqueness. The BoB representation would not distinguish between these two molecules. Only after addition of higher-order many-body potential terms (e.g., the 3-body Axilrod-Teller-Muto potential), the spurious degeneracy is lifted.

Based on this simple example, an important lesson learned is that collective effects which go beyond pairwise potentials are of vital importance for the accurate modeling of fundamental properties such as energies. While adhering to the ideas of bagging for efficiency, a representation consisting of extended bags can be constructed; each may contain interatomic interaction potentials up to three- and four-body terms. BAML was formulated in this way, where (1) all pairwise nuclear
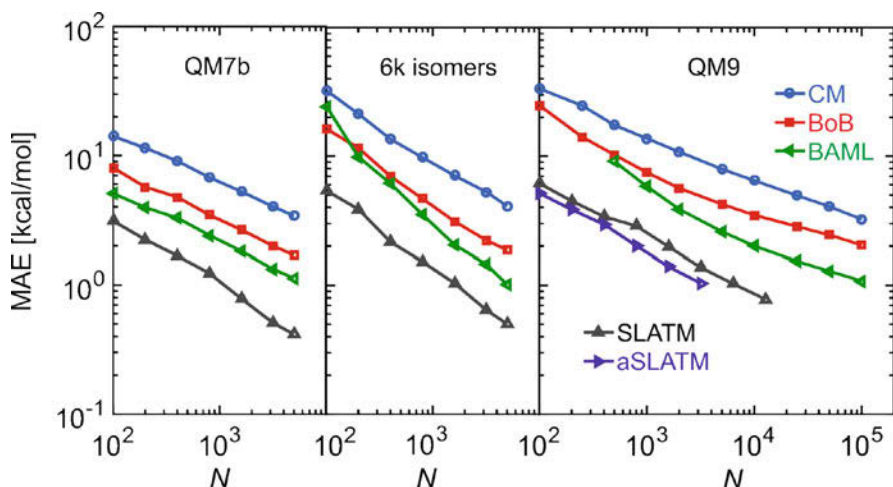
**Fig. 3** Comparison of the learning curves for different representations for three typical datasets (QM7b (Rupp et al. 2012), QM9 (Ramakrishnan et al. 2014; Ruddigkeit et al. 2012), and 6k isomers from QM9). Note that the size and composition for molecules in all the three datasets are comparable, i.e., the dimensionality $d$'s of these systems is similar; hence almost the same learning rates are observed for all representations with no (or less) suffer from uniqueness issue. For QM7b dataset, a much lower offset is shown as the relevant molecules are much more relaxed than those in QM9 and 6k isomers; thus given any representation, its target similarity is larger for this dataset compared to others

repulsions are replaced by Morse/Lennard-Jones potentials for bonded/nonbonded atoms, respectively, and (2) the inclusion of three- and four-body interactions of covalently bonded atoms is achieved using periodic angular and torsional terms, with their functional form and parameters extracted from the universal force field (UFF) (Huang and von Lilienfeld 2016; Rappe et al. 1992). BAML achieves a noticeable boost of performance when compared to BoB or CM. Interestingly, the performance is systematically improving upon inclusion of higher and higher-order many-body terms, as the proposed energy model is getting more and more realistic, i.e., increasing similarity to target. Meanwhile and not surprisingly, the uniqueness issue, existing in two-body representations such as BoB, is also resolved (see Fig. 3). The main drawback of BAML, however, is that it requires pre-existing force fields, implying a severe bias when it comes to new elements or bonding scenarios. It would therefore be desirable to identify a representation which is more compact and ab initio in nature.

The so-called SLATM representation (Huang and von Lilienfeld 2017) enjoys all these attributes. It has two variants: a local and a global one. The basic idea of SLATM is to represent an atom indexed $I$ in a molecule by accounting for all possible interactions between atom $I$ and its neighboring atoms through many-body potential terms multiplied by a normalized Gaussian distribution centered on the

relevant variable (distance or angle). So far, one-, two-, and three-body terms have been considered. The one-body term is simply represented by the nuclear charge, while the two-body part is expressed as

$$\frac{1}{2}Z_I \sum_{J \neq I} Z_J \delta(\mathbf{r} - \mathbf{R}_{IJ}) g(\mathbf{r}) \tag{35}$$

where $\delta(\cdot)$ is set to normalized Gaussian function $\delta(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2}$ and $g(r)$ is a distance-dependent scaling function, capturing the locality of chemical bond and chosen to correspond to the leading order term in the dissociative tail of the London potential $g(R) = \frac{1}{R^6}$. The three-body distribution reads

$$\frac{1}{3}Z_I \sum_{J \neq K \neq I} Z_J Z_K \delta(\theta - \theta_{IJK}) h(\theta, \mathbf{R}_{IJ}, \mathbf{R}_{IK}) \tag{36}$$

where $\theta$ is the angle spanned by vector $\mathbf{R}_{IJ}$ and $\mathbf{R}_{IK}$ (i.e.,$\theta_{IJK}$) and treated as a variable. $h(\theta, \mathbf{R}_{IJ}, \mathbf{R}_{IK})$ is the three-body contribution depending on both internuclear distance and angle and is chosen in form to model the Axilrod and Teller (1943) and Muto (1943) vdW potential:

$$h(\theta, \mathbf{R}_{IJ}, \mathbf{R}_{IK}) = \frac{1 + \cos\theta \cos\theta_{JKI} \cos\theta_{KIJ}}{(R_{IJ} R_{IK} R_{KJ})^3} \tag{37}$$

Now we can build the atomic version aSLATM for an atom $I$ through concatenation of all the different many-body potential spectra involving atom $I$ as displayed in Eqs. (35) and (36). As for the global version SLATM, it simply corresponds to the sum of the atomic spectra.

SLATM and aSLATM outperforms all other representations discussed so far, as evidenced by learning curves shown in Fig. 3. This outstanding performance is due to several aspects: (1) almost all the essential physics in the systems is covered, including the locality of chemical bonds as well as many-body dispersion; (2) the inclusion of 3-body terms significantly improves the learning; and (3) the spectral distribution of radial and angular feature now circumvents the problem of sorting within each feature bag, allowing for a more precise match of atomic environments.

Most recently, the FCHL representation has been introduced (Faber et al. 2018). It amounts to a radial distribution in elemental and structural degrees of freedom. The configurational degrees of freedom are expanded up to three-body interactions. Four-body interactions were tested but did not result in any additional improvements. For known datasets, FCHL-based QML models reach unprecedented predictive power and even outperform aSLATM and SOAP (see below). In the case of the QM9 dataset, for example, FCHL-based models of atomization energies reach chemical accuracy after training on merely ~1'000 molecules.

### 4.4.2  Density Expansion-Based Representation

Within the smooth overlap of atomic positions (SOAP) (Bartók et al. 2013) idea of a representation, an atom $I$ in a molecule is represented as the local density of atoms around $I$. Specifically, it is represented by a sum of Gaussian functions with variance $\sigma^2$ within the environment (including the central atom $I$ and its neighboring atoms $Q$'s), with the Gaussian functions centered on $Q$'s and $I$:

$$\rho_I(\mathbf{r}) = \sum_Q \exp\left(-\frac{(\mathbf{r} - \mathbf{R}_Q)^2}{2\sigma^2}\right) \tag{38}$$

where $\mathbf{r}$ is the vector from the central atom $I$ to any point in space, while $\mathbf{R}_Q$ is the vector from atom $I$ to its neighbor $Q$. The overlap of $\rho_I$ and $\rho_J$ then can be used to calculate a similarity between atoms $I$ and $J$. However, this similarity is not rotationally invariant. To overcome this, we can integrate out the rotational degrees of freedom for all three-dimensional rotations $\hat{R}$, and thus the SOAP kernel is defined:

$$\tilde{k}(I, J) = \int d\hat{R} \left| \int d\mathbf{r} \rho_I(\mathbf{r}) \rho_J(\hat{R}\mathbf{r}) \right|^2, \tag{39}$$

To enforce the self-similarity to be normalized, the final SOAP similarity measure takes the form of

$$k(I, J) = \frac{\tilde{k}(I, J)}{\sqrt{\tilde{k}(I, I)\tilde{k}(J, J)}} \tag{40}$$

The integration in Eq. (39) can be carried out by first expanding $\rho_I(\mathbf{r})$ in Eq. (38) in terms of a set of basis functions composed of orthogonal radial functions and spherical harmonics and then collecting the elements in the rotationally invariant power spectrum, based on which $k$ can be easily calculated. The interested reader is referred to Bartók et al. (2013).

SOAP has been used extensively and successfully to model systems such as silicon bulk or water clusters, each separately with many configurations. These elemental or binary systems are relatively simple as the diversity of chemistries encoded by the atomic environments is rather limited. A direct application of SOAP to molecules where there are substantially more possible atomic environments, however, yields learning curves with rather large offsets. This is not such a surprise, as essentially the capability of atomic densities to differentiate between different atom pairs, atom triples, and so on is not so great. This shortcoming remains even if one treats different atom pairs as different variables, as was adopted in De et al. (2016); averaging out all rotational degrees of freedom might also impede the learning progress due to loss of relevant information. To amend some of these problems, a special kernel, the RE-Match kernel (De et al. 2016), was

introduced. And most recently, combining SOAP with a multi-kernel expansion enabled additional improvements in predictive power (Bartók et al. 2017).

## 5     Training Set Selection

The last section of this chapter deals with the question of how to select training sets. The selection procedure can have a severe effect on the performance. The predictive accuracy appears to be very sensitive on how we sample the training molecules for any given representation (or better ones). Training set selection can actually be divided into two parts: (1) How to create training set. The general principle is that the training set should be representative, i.e., it follows the same distribution as all possible test molecules in terms of input and output. This will formally prevent extrapolation and thereby minimize prediction errors. (2) How to optimize the training set composition.

The majority of algorithms in literature deal with (2) assuming the existence of some large dataset (or a dataset trivial to generate) from which one can draw using algorithms such as ensemble learning, genetic evolution, or other "active learning"-based procedures (Podryabinkin and Shapeev 2017). All of these methods have in common that they select the training set from a given set of configurations based only on the unlabeled data. This is particularly useful for "learning on the fly"-based ab initio molecular dynamics simulations Csányi et al. (2004), where expensive quantum-mechanical calculation is carried out only when the configurations are sufficiently "new."

Step 1 stands out as a challenging task and few algorithms are competent. The most ideal approach is of course an algorithm that can do both parts within one step; the only competent method we know is the "amons" approach. We will elaborate on all these concepts below.

## 5.1     Genetic Optimization

To the best of our knowledge, the first application of a GA for generation and study of optimal training set compositions for QML model was published in Browning et al. (2017). The central idea of this approach is outlined as follows. For a given set ($S_0$) containing overall $N$ molecules, the GA procedure consists of three consecutive steps to obtain the "near-optimal" subset of molecules from $S_0$ for training the ML model (Browning et al. 2017): (a) Randomly choose $N_1$ molecules as a trial training set $s_1$; repeat $M$ times. This forms a population of training sets, termed the parent population and labeled as $\hat{s}^{(1)} = \{s_1, s_2, \ldots, s_M\}$. (b) An ML model is trained on each $s_i$ and then tested on a fixed set of out-of-sample molecules, resulting in a mean prediction error $e_i$, which is assigned to $s_i$ as a measure of how fit $s_i$ is as the "near-optimal" training set and dubbed "fitness." Therefore, the smaller $e_i$ is, the larger the fitness is. (c) $\hat{s}^{(1)}$ is consecutively evolved through selection (to determine which $s_i$'s in $\hat{s}^{(1)}$ should remain in the population to produce a temporarily refined

smaller set $\hat{t}^{(1)}$; a set $s_i$ with larger fitness means higher probability to be kept in $\hat{t}^{(1)}$), crossover (to update $\hat{s}^{(1)}$ from $\hat{t}^{(1)}$ and the new $\hat{s}$ is labeled as $\hat{s}^{(2)}$ with each set $s_i$ in $\hat{s}^{(2)}$ obtained through mixing the molecules from two $s_i$'s in $\hat{t}^{(1)}$), and mutation (to change molecules in some $s_i$'s in $\hat{s}^{(2)}$ randomly to promote diversity in $\hat{s}^{(2)}$, e.g., replace -$CH_2$- fragment by -NH- for some molecule. (d) Go to step (b) and repeat the process until there is no more change in the population and the fitness ceases to improve . We label the final updated trial training set as $\hat{s}$.

It's obvious that the molecules in $\hat{s}$ should be able to represent all the typical chemistry in all molecules in $S_0$, such as linear, ring, cage-like structure, and typical hybridization states ($sp, sp^2, sp^3$) if they are abundant in $S_0$. Once trained on $\hat{s}$, the ML model is guaranteed to yield typically significantly better results as the fitness is constantly increasing. This is not useful since the GA "tried" this already; the usefulness has to be assessed by the generalizability of $\hat{s}$ as training set to test on a new set of molecules is not seen in $S_0$. Indeed, as shown in Browning et al. (2017), significant improvements in offsets can be obtained when compared to random sampling. While the remaining out-of-sample error is still substantial, this is not surprising due to the use of less advantageous representations. One of the key findings in this study were that upon genetic optimization, (i) the distance distributions between training molecules were shifted outward and (ii) the property distributions of training molecules were fattened.

## 5.2    Amons

We note that the naive application of active learning algorithms will still result in QML models which suffer from lack of transferability, in particular when it comes to the prediction of larger compounds or molecules containing chemistries not present in the training set. Due to the size of chemical compound space, this issue still imposes a severe limitation for the general applicability of QML. These problems can, at least partially, be overcome by exploring and exploiting the locality of an atom in molecule (Huang and von Lilienfeld 2017), resulting from the nearsightedness principle in electronic systems (Prodan and Kohn 2005; Fias et al. 2017).

We consider a valence saturated query molecule for illustration, for which we try to build an "ideal" training set. As is well known, any atom $I$ (let us assume a $sp^3$ hybridized C) in the molecule is characterized by itself and its local chemical environment. To a first-order approximation, we may consider its coordination number ($CN$ for short) to be a distinguishing measure of its atomic environment, and we can roughly say that any other carbon atom with a coordination number of 4 is similar to atom $I$, as their valence hybridization states are all $sp^3$. Another carbon atom with $CN = 3$ in hybridization state of $sp^2$ would be significantly different compared to atom $I$. It is clear, however, that $CN$ as an identifier of atomic environment type is not enough: An $sp^3$ hybridized C atom in methane molecule (hereafter we term it as a genuine C-$sp^3$ environment) is almost purely

**Fig. 4** Comparison of the learning curves obtained for one molecule (see the inset) from random selection of training set and amons, respectively. For each red scatter point, errors were averaged over 100 random samplings
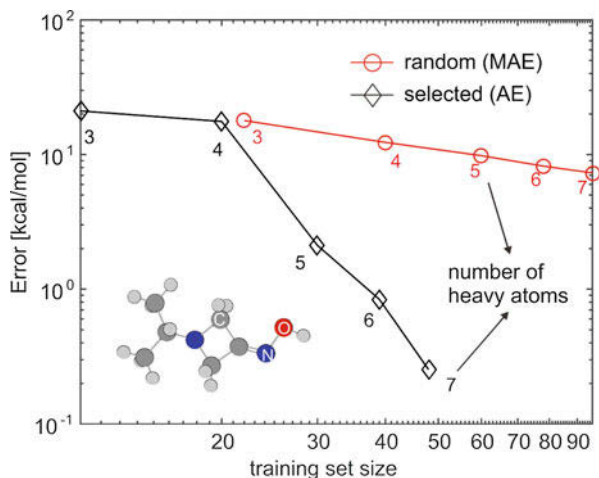
covalently bonded to its neighbors, while in $CH_3OH$, noticeable contributions from ionic configurations appear in the valence bond wavefunction due to the significant electronegativity difference between C and O atoms. Thus one would expect very different atomic properties for the $sp^3$-C atoms in these two environments as manifested, for instance, in their atomic energy, charge, or $^{13}$C-NMR shift. Alternatively, we can say that oxygen as a neighboring atom to $I$ has perturbed the ideal $sp^3$ hybridized C to a much larger extent in $CH_3OH$ than the H atom has in methane. To account for these differences, we can simply include fragments which contain $I$ as well as all its neighbors. Thus we can obtain a set of fragments, for each of which the bond path between $I$ and any other atom is 1.

Extending this kind of reasoning to the second neighbor shell, we can add new atoms with a bond path of 2 relative to atom $I$ in order to account for further, albeit weaker, perturbation to atom $I$. As such, we can gradually increase the size of included fragments (characterized by the number of heavy atoms) until we believe that all effects on atom $I$ have been accommodated. The set of unique fragments can then be used as a training set for a fragment-based QML model. Note that we saturate all fragments by hydrogen atoms. These fragments can be regarded as effective quasi-atoms which are defined as a̲tom in mo̲lecule, or "amon." Since amons repeat throughout chemical space, they can be seen as the "words" of chemistry (target molecules being "sentences") or as "DNA" of chemistry (target molecules being genes and properties their function). Given the complete set of amons, any specific, substantially larger, query molecule can be queried. Used in conjunction with an atomic representation such as aSLATM or FCHL, amons enable a kind of chemical extrapolation which holds great promise to more faithfully and more efficiently explore vast domain chemical space (Huang and von Lilienfeld 2017).

To demonstrate the power of amons, we show the example of predicting the potential energy of a molecule present as an inset in Fig. 4. With amons as the

training set, chemical accuracy (1 kcal/mol) is reached after training on only 40 amons (with amons being not larger than 6 heavy atoms). Sampling amons at random, the slope of the learning curve is substantially worse.

# 6    Conclusion

We have discussed primarily the basic mathematical formulations of all typical ingredients of quantum machine learning (QML) models which can be used in the context of quantum-mechanical training and testing data. We explained and reviewed why ML models can be fast and accurate when predicting quantum-mechanical observables for out-of-sample compounds. It is the authors' opinion that QML can be seen as a very promising approach, enabling the exploration of systems and problems which hitherto were not amenable to traditional computational chemistry methods.

In spite of the significant progress made within the last few years, the field QML is still very much in a stage of infancy. This should be clear when considering that the properties that have been explored so far are rather limited and relatively fundamental. The primary focus has been on ground state or local minimum properties. Application to excited states still remains a challenge (Ramakrishnan et al. 2015b), just as well as conductivity, magnetic properties, or phase transitions. We believe that new and efficient representations will have to be developed which properly account for all the relevant degrees of freedom at hand.

# References

Axilrod BM, Teller E (1943) Interaction of the van der Waals type between three atoms. J Chem Phys 11(6):299–300. https://doi.org/10.1063/1.1723844, http://scitation.aip.org/content/aip/journal/jcp/11/6/10.1063/1.1723844

Bader RF (1990) Atoms in molecules: a quantum theory. Clarendon Press, Oxford

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403. https://doi.org/10.1103/PhysRevLett.104.136403

Bartók AP, Kondor R, Csányi G (2013) On representing chemical environments. Phys Rev B 87:184115. https://doi.org/10.1103/PhysRevB.87.184115

Bartók AP, De S, Poelking C, Bernstein N, Kermode JR, Csányi G, Ceriotti M (2017) Machine learning unifies the modeling of materials and molecules. Sci Adv 3(12):e1701816 https://doi.org/10.1126/sciadv.1701816, http://advances.sciencemag.org/content/3/12/e1701816

Browning NJ, Ramakrishnan R, von Lilienfeld OA, Roethlisberger U (2017) Genetic optimization of training sets for improved machine learning models of molecular properties. J Phys Chem Lett 8(7):1351. https://doi.org/10.1021/acs.jpclett.7b00038

Csányi G, Albaret T, Payne MC, Vita AD (2004) "Learn on the fly": a hybrid classical and quantum-mechanical molecular dynamics simulation. Phys Rev Lett 93:175503

De S, Bartok AP, Csanyi G, Ceriotti M (2016) Comparing molecules and solids across structural and alchemical space. Phys Chem Chem Phys 18:13754–13769. https://doi.org/10.1039/C6CP00415F

Faber FA, Christensen AS, Huang B, von Lilienfeld OA (2018) Alchemical and structural distribution based representation for improved QML. J Chem Phys 148:241717. https://doi.org/10.1063/1.5020710. arXiv preprint arXiv:171208417

Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, Vinyals O, Kearnes S, Riley PF, von Lilienfeld OA (2017) Machine learning prediction errors better than DFT accuracy. arXiv preprint arXiv:1702.05532. https://arxiv.org/abs/1702.05532

Fasshauer G, McCourt M (2016) Kernel-based approximation methods using MATLAB. World Scientific, Singapore

Fias S, Heidar-Zadeh F, Geerlings P, Ayers PW (2017) Chemical transferability of functional groups follows from the nearsightedness of electronic matter. Proc Natl Acad Sci 114(44):11633–11638

Ghiringhelli LM, Vybiral J, Levchenko SV, Draxl C, Scheffler M (2015) Big data of materials science: critical role of the descriptor. Phys Rev Lett 114:105503. https://doi.org/10.1103/PhysRevLett.114.105503

Greeley J, Jaramillo TF, Bonde J, Chorkendorff I, Nørskov JK (2006) Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. Nat Mater 5(11):909–913

Hansen K, Biegler F, von Lilienfeld OA, Müller KR, Tkatchenko A (2015) Interaction potentials in molecules and non-local information in chemical space. J Phys Chem Lett 6:2326. https://doi.org/10.1021/acs.jpclett.5b00831

Huang B, von Lilienfeld OA (2016) Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity. J Chem Phys 145(16):161102. https://doi.org/10.1063/1.4964627

Huang B, von Lilienfeld OA (2017) Chemical space exploration with molecular genes and machine learning. arXiv preprint arXiv:170704146

Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. Biometrika 87(1):1–13

Kirkpatrick P, Ellis C (2004) Chemical space. Nature 432(7019):823–823. https://doi.org/10.1038/432823a

Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Fragment molecular orbital method: an approximate computational method for large molecules. Chem Phys Lett 313(3–4):701–706. https://doi.org/10.1016/S0009-2614(99)00874-X

Muto Y (1943) Force between nonpolar molecules. J Phys Math Soc Jpn 17:629–631

Pilania G, Gubernatis JE, Lookman T (2017) Multi-fidelity machine learning models for accurate bandgap predictions of solids. Comput Mater Sci 129:156–163

Podryabinkin EV, Shapeev AV (2017) Active learning of linearly parametrized interatomic potentials. Comput Mater Sci 140:171–180

Prodan E, Kohn W (2005) Nearsightedness of electronic matter. Proc Natl Acad Sci USA 102(33):11635–11638. https://doi.org/10.1073/pnas.0505436102

Ramakrishnan R, von Lilienfeld OA (2015) Many molecular properties from one kernel in chemical space. Chimia 69(4):182. https://doi.org/10.2533/chimia.2015.182, http://www.ingentaconnect.com/content/scs/chimia/2015/00000069/00000004/art00005

Ramakrishnan R, Dral P, Rupp M, von Lilienfeld OA (2014) Quantum chemistry structures and properties of 134 kilo molecules. Sci Data 1:140022. https://doi.org/10.1038/sdata.2014.22

Ramakrishnan R, Dral P, Rupp M, von Lilienfeld OA (2015a) Big data meets quantum chemistry approximations: the Δ-machine learning approach. J Chem Theory Comput 11:2087–2096. https://doi.org/10.1021/acs.jctc.5b00099

Ramakrishnan R, Hartmann M, Tapavicza E, von Lilienfeld OA (2015b) Electronic spectra from TDDFT and machine learning in chemical space. J Chem Phys 143:084111. http://arxiv.org/abs/1504.01966

Rappe AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc 114(25):10024–10035. https://doi.org/10.1021/ja00051a040

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. Adaptative computation and machine learning series. University Press Group Limited. https://books.google.ch/books?id=vWtwQgAACAAJ

Ruddigkeit L, van Deursen R, Blum LC, Reymond JL (2012) Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. J Chem Inf Model 52(11): 2864–2875. https://doi.org/10.1021/ci300415d

Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA (2012) Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 108(5):058301. https://doi.org/10.1103/PhysRevLett.108.058301

Samuel AL (2000) Some studies in machine learning using the game of checkers. IBM J Res Dev 44(1.2):206–226

Todeschini R, Consonni V (2008) Handbook of molecular descriptors, vol 11. Wiley, Weinheim

von Lilienfeld OA (2018) Quantum machine learning in chemical compound space. Angew Chemie Int Ed 57(16):4164–4169. https://doi.org/10.1002/anie.201709686

von Lilienfeld OA, Ramakrishnan R, Rupp M, Knoll A (2015) Fourier series of atomic radial distribution functions: a molecular fingerprint for machine learning models of quantum chemical properties. Int J Quantum Chem 115(16):1084–1093. https://doi.org/10.1002/qua.24912

# Machine Learning of Atomic-Scale Properties Based on Physical Principles

# 87

Michele Ceriotti, Michael J. Willatt, and Gábor Csányi

## Contents

## Abstract

We briefly summarize the kernel regression approach, as used recently in materials modeling, to fitting functions, particularly potential energy surfaces, and highlight how the linear algebra framework can be used to both predict and

M. Ceriotti · M. J. Willatt
Laboratory of Computational Science and Modelling, Institute of Materials, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: michele.ceriotti@epfl.ch

G. Csányi (✉)
Engineering Laboratory, University of Cambridge, Cambridge, UK
e-mail: gc121@cam.ac.uk

train from linear functionals of the potential energy, such as the total energy and atomic forces. We then give a detailed account of the smooth overlap of atomic position (SOAP) descriptor and kernel, showing how it arises from an abstract representation of smooth atomic densities and how it is related to several popular density-based descriptors of atomic structure. We also discuss recent generalizations that allow fine control of correlations between different atomic species, prediction, and fitting of tensorial properties and also how to construct structural kernels – applicable to comparing entire molecules or periodic systems – that go beyond an additive combination of local environments.

# 1    Introduction

There has been a surge of activity during the last couple of years in applying machine learning methods to materials and molecular modeling problems that was largely fueled by the evident success of these techniques in what can loosely be called artificial intelligence. These successes have followed from the collective experience that the scientific community has gained in fitting high volumes of data with very complex functional forms that involve a large number of free parameters while still keeping control of the regularity and thus avoiding catastrophic over-fitting. In the context of molecular modeling, empirical fitting of potential energy surfaces has of course been used for many decades. Indeed it is generally held that this is the only practical way to simulate very large systems (many thousands of atoms) over long time scales (millions of time steps) (Finnis 2004).

Traditionally, when fitting empirical models of atomic interactions, regularity was ensured by writing functional forms that are expressed in terms of one-dimensional functions, e.g., pair potentials, spherically symmetric atomic electron densities, bond orders (as a function of number of neighbors), etc. Such functions are easy to inspect visually to ensure that they are physically and chemically meaningful, e.g., that pair potentials go to zero at large distances and are strongly repulsive at close approach, that atomic electron densities are decreasing with distance, that electron density embedding functions are convex, etc. Moreover, these natural properties are easy to build into the one-dimensional functional forms or enforced as constraints in the parameter optimization. It is widely held that employing such "physically meaningful" functional forms is the key to achieving good *transferability* of the empirical models (Brenner 2000).

It is also recognized, however, that the limited functional forms that can be built from these one-dimensional functions ultimately limit the accuracy that these empirical models can achieve. In trying to replace them by high-dimensional fits using much more flexible functional forms, two things immediately have to change. The first is the target data. When fitting only a few parameters, it is natural to demand that important observables that are deemed to be central to the scientific questions being addressed are reproduced correctly, and it is easiest to do this if they are part of the fit, e.g., melting points and other phase boundaries, radial distribution functions, etc. But in the case of very many parameters, their optimization also takes a significant

number of evaluations, and it becomes impractical to use complex observables as targets. Moreover, there is a drive toward using a "first principles" approach, i.e., that the potentials should actually reproduce the real Born-Oppenheimer potential energy surface with sufficient accuracy *and therefore* the scientifically relevant observables also. So it is natural to take as target data values of the potential energy surface, as computed by the electronic structure method of choice. The large number of free parameters can then easily be counterbalanced by a large amount of calculated input data.

The second thing that has to change is how the smooth physically meaningful behavior of the potential is controlled. It is not practical to inspect manually high-dimensional functions to ensure that their predictions are physically and chemically meaningful for all possible configurations. Therefore it becomes even more important to build into the functional forms as much prior information as possible about limiting behavior and *regularity* (the technical word for the kind of smoothness we are interested in). Reviewing the recent work, this paper sets out an example framework for how to do this. The key goals are to create functional forms that preserve the (i) invariance of the properties over permutation of like atoms, (ii) invariance of scalar and covariance of tensorial properties with three-dimensional rotations, and (iii) continuity and regularity with respect to changes in atomic coordinates, including compact support of atomic interactions by including finite cutoffs.

Evidence is accumulating showing that enforcing these physically motivated properties strictly is enormously beneficial, and several of the most successful machine learning schemes for atomic-scale systems are built around symmetry arguments. One possible approach is to describe the system in terms of internal coordinates that satisfy automatically rotational invariance and then symmetrize explicitly the vector of descriptors or the functional relation between the descriptors and the properties. Permutationally invariant polynomials are an example that has been very effective to model the potential energy surfaces of small molecules (see, e.g., the work of Braams and Bowman 2009). Sorting the descriptors according to interatomic distances has also been used as a way of obtaining permutation invariance at the cost of introducing derivative discontinuities (Rupp et al. 2012; Faber et al. 2015; Zhang et al. 2018). Another possibility, which we will focus on in this paper, starts from a representation of each structure in terms of atomic densities – which are naturally invariant to atom permutations – and then builds a descriptor that is further invariant to translations and rotations also.

Either way, once an appropriate description of each structure has been obtained, further regularization can be achieved at the level of the regression scheme. To this end, two prominent techniques are the use of artificial neural networks and kernel ridge regression (Bishop 2016). We use the latter formalism here, and many further details about these techniques can be found in the rest of this volume. The kernel approach starts with the definition of a kernel function, which will be combined with a set of representative atomic configurations to construct basis functions for the fit. It is a scalar function – at least when learning scalar quantities – with two input arguments and in the present case two atomic structures. Its value should quantify

the similarity of the atomic configurations represented by its two arguments, and it can (but does not have to) be defined starting from their associated descriptors. The value should be the largest when its two arguments are equal (or equivalent up to symmetry operations) and smallest for maximally different configurations. The degree to which the kernel is able to capture the variation of the function when varying the atomic configuration will determine how efficient the fit is. The better the correspondence, the fewer basis functions that are needed to achieve a given accuracy of fit.

## 2    Kernel Fitting

We start by giving a concise account of the kernel regression fitting approach, for more details see Bishop (2016), Rasmussen and Williams (2006), and Schölkopf and Smola (2002). A function defined on an atomic structure is represented as a linear sum over kernel basis functions,

$$f(\mathcal{A}) = \sum_{\mathcal{B} \in M} x_{\mathcal{B}} K(\mathcal{A}, \mathcal{B}), \tag{1}$$

where the sum runs over a *representative set* of configurations $M$, selected from the total set $N$ of input configurations. The set of coefficients, combined into a vector $\mathbf{x}$, are determined by solving the linear system that is obtained when the available data (e.g., values of the target function evaluated for a set of structures) are substituted into Eq. (1). In the simplest case, there is one input data value corresponding to each atomic configuration. Let $\mathbf{y}$ be the vector of all available input data and $\mathbf{K}$ be the kernel matrix with rows and columns corresponding to atomic structures, so that the element of $\mathbf{K}$ with row and column corresponding to structures $\mathcal{A}$ and $\mathcal{B}$, respectively, is $K(\mathcal{A}, \mathcal{B})$. The fit is then obtained by solving a linear system in the least squares sense, i.e., minimizing the quadratic loss function:

$$\ell(\mathbf{x}) = \|\mathbf{K}\mathbf{x} - \mathbf{y}\|^2. \tag{2}$$

The textbook case is when the set of all configurations for which we have target data available is used in its entirety as the representative set (i.e., $N = M$), $\mathbf{K}$ is square, and as long as it is invertible, the optimal solution is:

$$\mathbf{x} = \mathbf{K}^{-1}\mathbf{y}. \tag{3}$$

In practice, for large data sets, using all the configurations in the data set as representatives is unnecessary. In this case, $M \subset N$, the solution is given by the pseudoinverse,

$$\mathbf{x}_M = (\mathbf{K}_{MN}\mathbf{K}_{NM})^{-1}\mathbf{K}_{MN}\mathbf{y}_N, \tag{4}$$

where we used subscripts to emphasize the set that the vector elements correspond to, e.g., $\mathbf{y} \equiv \mathbf{y}_N$ is the data vector with one element for each input data structure and $\mathbf{x} \equiv \mathbf{x}_M$ is the vector of coefficients, one for each representative configuration. The subscripts on the kernel matrix denote array slices, i.e., $\mathbf{K}_{MN} = \mathbf{K}_{NM}^{\top}$ is the rectangular matrix whose elements correspond to the kernel values between the representative configurations and the input configurations.

Using a representative set much smaller than the total number of structures has significant advantages in terms of computational cost, often with no reduction in fitting accuracy. The training cost is dominated by computing the pseudoinverse, which scales as $O(NM^2)$, which is linear in the size of the training data, $N$, evaluating the model scales as $O(M)$, now independent of the size of the training data. These cost scalings are analogous to those of artificial neural networks with a fixed number of nodes.

While the above solutions are formally correct, it is widely recognized that they lead to numerical instability and *overfitting*, i.e., they are solutions that attempt to maximize the fit to the input data, even when this might not be desirable, which is almost always the case. At first sight, this might sound surprising, since electronic structure calculations can be made deterministic, with precise convergence behavior in terms of its parameters, such as k-point sampling, SCF tolerance, etc. However, practical calculations are never converged to machine precision, and the resulting inconsistencies between the potential energy values for different configurations are not something that is desirable to propagate to a fitted potential energy surface. The magnitude of such inconsistencies can be easily assessed before the fit is made. Previous experience (Szlachta et al. 2014; Dragoni et al. 2018) suggests that in large databases for materials applications, the error due to k-point sampling is dominant and difficult to reduce below a meV/atom due to the associated computational cost.

In case we are fitting a potential energy surface with a representation that does not characterize the atomic positions of the whole system completely due to, e.g., a finite cutoff, or some other choices made to gain computational efficiency, the fit is not expected to be exact, irrespective of the amount of input data. Sometimes, such *model error* can also be assessed a priori, e.g., in the case of a finite cutoff by measuring the contribution made to forces on an atom by other atoms beyond the cutoff (Bernstein et al. 2009; Deringer and Csányi 2017; Fujikake et al. 2018).

These two considerations suggest that allowing some "looseness" in the linear system might be beneficial, because it can be exploited to allow smaller linear coefficients, making the fit more regular and thus better at extrapolation. We collect the errors we expect in the fit of each target data value on the diagonal of an $N \times N$ matrix, $\mathbf{\Lambda}$. The common procedure to regularizing the problem is due to Tikhonov et al. (1995). Specifically, in "kernel ridge regression" (and the equivalent "Gaussian process regression," a Bayesian view of the same), the Tikhonov matrix is chosen to be the kernel matrix between the $M$ representative points, $\mathbf{K}_{MM}$. With highly regular ("smooth") kernel functions, this regularization leads to smooth fits, and the sizes of the elements of $\mathbf{\Lambda}$ control the trade-off between the accuracy of the fit and smoothness. The corresponding solutions are

$$\mathbf{x} = (\mathbf{K} + \mathbf{\Lambda})^{-1}\mathbf{y}. \tag{5}$$

for the square problem, and

$$\mathbf{x}_M = (\mathbf{K}_{MM} + \mathbf{K}_{MN}\mathbf{\Lambda}^{-1}\mathbf{K}_{NM})^{-1}\mathbf{K}_{MN}\mathbf{\Lambda}^{-1}\mathbf{y}_N, \tag{6}$$

for the rectangular problem, where we again emphasized the index sets. This solution is equivalent to minimizing

$$\|\mathbf{K}\mathbf{x} - \mathbf{y}\|_{\mathbf{\Lambda}^{-1}}^2 + \|\mathbf{x}\|_{\mathbf{K}}^2, \tag{7}$$

which shows that the inverse of the tolerances in $\mathbf{\Lambda}$ are equivalent to weights on the different data points.

With the solution of the linear system in hand, the value of the fitted function for a new structure $\mathcal{C}$ can be written as

$$f(\mathcal{C}) = \mathbf{K}_{\mathcal{C}M}\mathbf{x}_M. \tag{8}$$

Note that the $\mathbf{K}_{\mathcal{C}M}$ slice is just a vector, with elements given by the kernel between the new structure $\mathcal{C}$ and the structures in the representative set $M$.

## 2.1 Selection of a Representative Set

Next we describe some ways to choose the set of representative environments over which the sum in Eq. (1) is taken. This can be done by simple random sampling, but we find it advantageous to use this freedom to optimize interpolation accuracy. Among the many strategies that have been proposed to this end (Hartigan and Wong 1979; Prabhakaran et al. 2012), we discuss two that have been used successfully in this context. One approach to this is to maximize the dissimilarity between the elements of the representative set. A greedy algorithm to select the configurations for the representative set is "farthest point sampling," in which we start with a randomly selected structure and then iteratively pick as the next structure the one which is farthest away from any of the structures already in the set (Gonzalez 1985; Ceriotti et al. 2013; Bartók et al. 2017). The distance between two structures is measured in the "kernel metric," defined as (Schölkopf and Smola 2002):

$$d^2(\mathcal{A}, \mathcal{B}) = K(\mathcal{A}, \mathcal{A}) + K(\mathcal{B}, \mathcal{B}) - 2K(\mathcal{A}, \mathcal{B}). \tag{9}$$

This algorithm performed well for selecting molecules in regression tasks, enabling the significant reduction of the data set sizes for a given level of accuracy (De et al. 2016).

Another technique that has been successfully used is based on matrix factorization, which is particularly appealing when the kernel function is linear or a low-order

polynomial of the descriptor vector. Consider the matrix of descriptor vectors, $\mathbf{D}$, in which each row is the descriptor vector of an input atomic configuration, such that a linear kernel is $\mathbf{K} = \mathbf{D}\mathbf{D}^\top$. We are looking to select rows, many fewer than the total number, which span as much of the space as all rows span. This is a problem of matrix representation, specifically that the representative set should serve as a low-rank approximation of $\mathbf{K}$ and/or $\mathbf{D}$. One solution to this is called CUR matrix decomposition (Mahoney and Drineas 2009), which can be applied to either $\mathbf{K}$ or $\mathbf{D}$, the latter being much cheaper if the length of the descriptor vectors is less than the number of data points.

To determine the optimal set of representative configurations, we start with a singular value decomposition of $\mathbf{D}$,

$$\mathbf{D} = USV^\top. \tag{10}$$

For each data point, a *leverage score* is calculated, essentially the weight that the top singular vectors have on that configuration.

$$\pi_{\mathcal{A}} = \frac{1}{k}\sum_{\xi=1}^{k}\left(u_{\mathcal{A}}^{\xi}\right)^2 \tag{11}$$

where $u_{\mathcal{A}}^{\xi}$ is the element of the $\xi$-th left singular vector that corresponds to structure $\mathcal{A}$. The sum runs over the first $k$ singular vectors, e.g., $k = 20$ is typical. The configuration $\mathcal{A}$ is included in the representative set with a probability that is proportional to its leverage score, $\pi_{\mathcal{A}}$. A deterministic variant is to select one structure $\mathcal{A}$ at a time – the one with the highest leverage score – delete the associated row from the descriptor matrix, and orthogonalize the remaining rows of $\mathbf{D}$ relative to it. The next data point can then be selected repeating the same procedure on the now smaller descriptor matrix (Imbalzano et al. 2018).

Note that in the Gaussian process literature, using a subset of the data to construct the basis is called *sparsification* (Quinonero-Candela and Rasmussen 2005; Snelson and Ghahramani 2006), even though the approximation relies on a low-rank matrix reconstruction rather than the kernel matrix being sparse.

## 2.2     Linear Combination of Kernels

When fitting interatomic potentials for materials, a model is constructed for the *atomic energy*, sometimes called the "site energy." This is both for computational efficiency and to reduce the complexity of the functional relation between structures and properties: each atomic energy is only a function of a limited number of degrees of freedom corresponding to the coordinates of the neighboring atoms and can therefore be evaluated independently from any other atomic energy. In fact this is the defining characteristic of an interatomic potential, in contrast to a quantum mechanical model that explicitly includes delocalized electrons. Going from atomic

energies to the total energy is trivial, the latter being the sum of the former. However, going in the other direction is not unambiguous. The total energy can be calculated from a quantum mechanical model, but the atomic energies are not defined uniquely, and it becomes part of the fitting task to find the best possible separation of the total energy into atomic energies. Treating these two transformations on the same footing helps. Suppose we want to predict the sum of function values for two (or more) configurations. For the simple case of the sum of two energies for structures $\mathcal{A}$ and $\mathcal{B}$, the prediction is, trivially, just the sum of the individual function value predictions, e.g.:

$$E_{\text{tot}} = E(\mathcal{A}) + E(\mathcal{B}) = \mathbf{K}_{\mathcal{A}M}\mathbf{x}_M + \mathbf{K}_{\mathcal{B}M}\mathbf{x}_M. \tag{12}$$

If we define a new "sum-kernel" to be the sum of kernel values between a number of new configurations and the representative set, the expression for the above total energy prediction takes the same form as the prediction of the individual function values. For some set $I$ of new configurations, let

$$^{\Sigma}\mathbf{K}_M = \sum_{\mathcal{A} \in I} \mathbf{K}_{\mathcal{A}M}, \tag{13}$$

where $^{\Sigma}\mathbf{K}_M$ is the *vector* of sum-kernel values, each element of which is the sum of the kernel between all the configurations in $I$ and a given configuration in the representative set $M$. The predicted total energy of the configurations in $I$ is then

$$E_{\text{tot}} = {}^{\Sigma}\mathbf{K}_M\mathbf{x}_M. \tag{14}$$

This same sum-kernel can be used to fit the model to sum data, rather than to individual function values. This is critical for the case of fitting interatomic potentials for materials systems, since only total energies, and not the atomic energies themselves, are available from electronic structure calculations. At the same time, in order to enforce a finite short range in the interatomic potential, we must express the potential as an atomic energy. Using the sum-kernel, this is straightforward, the original functional form in (1) can be retained, and then we now minimize:

$$\|{}^{\Sigma}\mathbf{K}\mathbf{x} - \mathbf{E}_{\text{tot}}\|^2, \tag{15}$$

where $^{\Sigma}\mathbf{K}$ is a matrix containing the sum-kernel values for all configurations in the input database and the representative set and the vector $\mathbf{E}_{\text{tot}}$ is the collection of corresponding total energy data.

## 2.3    Derivatives

The explicit analytic functional form of Eq. (1) leads to analytic derivatives with respect to the atomic coordinates, e.g., forces in the case of fitting an energy. Considering for the moment the simpler case in which we are computing the derivatives of an atom-centered quantity $f(\mathcal{A})$, we define $\nabla_{\mathcal{A}}$ as the vector of derivatives with respect to all the atomic coordinates in structure $\mathcal{A}$. We use the notation $\overleftarrow{\nabla}$ to indicate a derivative operator that applies to the first argument of the kernel and $\overrightarrow{\nabla}$ to indicate a derivative that applies to the second argument. The derivatives of $f(\mathcal{A})$ are nonzero only for atoms that belong to the structure $\mathcal{A}$ and are then given by differentiating Eq. (1):

$$\nabla_{\mathcal{A}} f(\mathcal{A}) = \sum_{\mathcal{B} \in M} x_{\mathcal{B}} \overleftarrow{\nabla}_{\mathcal{A}} K(\mathcal{A}, \mathcal{B}) = \mathbf{K}_{\nabla \mathcal{A} M} \mathbf{x}_M, \tag{16}$$

where we introduced the notation $\mathbf{K}_{\nabla \mathcal{A} M}$ to indicate the matrix that contains the derivatives of the kernels relative to all the relevant atomic coordinates. Similarly to the case of sums above, the gradient of the kernel function can also be used for fitting the model not to target values, but to *gradient data* (Solak et al. 2003). This is especially useful when the target represents a potential energy surface. When using typical electronic structure methods, the cost of computing the gradient with respect to all atomic positions is only a little bit more than the cost of computing the energy but yields much more information, $3n$ pieces of data for an $n$-atom structure. There are two approaches one can take to incorporate gradient information. In the first one, used in Bartók et al. (2010) and subsequent work of that group (Bartók et al. 2013a, 2017; Szlachta et al. 2014; Bartók and Csányi 2015; Deringer and Csányi 2017; John and Csányi 2017; Fujikake et al. 2018; Dragoni et al. 2018; Deringer et al. 2018; Caro et al. 2018; Rowe et al. 2018; Nguyen et al. 2018), the functional form for the energy is again retained to be the same as in Eq. (1). The corresponding loss function is

$$\| \mathbf{K}_{\nabla N M} \mathbf{x}_M - \mathbf{y}_{\nabla N} \|^2, \tag{17}$$

where $\mathbf{y}_{\nabla N}$ refers to the concatenated vector of gradients on all atoms in the set of input structures and $\mathbf{K}_{\nabla N M}$ to the corresponding matrix of kernel derivatives. The form of the solution for the coefficients is unchanged from (5) or (6) with $\mathbf{K}_{\nabla N M}$ taking the role of $\mathbf{K}_{N M}$.

In the second approach, used recently in Chmiela et al. (2017), derivatives of the kernel are the basis functions in the functional form of the fit,

$$f(\mathcal{A}) = \sum_{\mathcal{B} \in M} \mathbf{x}_{\nabla \mathcal{B}} \cdot \overrightarrow{\nabla}_{\mathcal{B}} K(\mathcal{A}, \mathcal{B}), \tag{18}$$

where $\mathbf{x}_{\nabla B}$ contains one weight for each of the derivatives relative to the atoms in structure $B$. The number of basis functions and corresponding coefficients is now much larger, $3nM$, for $n$-atom structures. Since the model is fitted to the derivatives, given by gradients of (18), the loss is

$$\|\mathbf{K}_{\nabla N \nabla M}\mathbf{x}_{\nabla M} - \mathbf{y}_{\nabla N}\|^2, \tag{19}$$

the target properties can be computed as

$$f(\mathcal{A}) = \mathbf{K}_{\mathcal{A}\nabla M}\mathbf{x}_{\nabla M}, \tag{20}$$

and their derivatives as

$$\nabla_{\mathcal{A}}f(\mathcal{A}) = \mathbf{K}_{\nabla \mathcal{A}\nabla M}\mathbf{x}_{\nabla M}. \tag{21}$$

The original motivation for this approach is apparent from Eq. (19) in which the matrix can be understood as a kernel directly between atomic forces (and in case of $M = N$, between the input data forces).

Both approaches constitute valid ways of learning a function from data representing its gradients, differing only in the choice of the kernel basis. The kernel-derivative basis functions could also be used in conjunction with a reduced representative set, and it is not yet clear which approach is better or indeed a combination: one could choose different basis functions (kernels or their derivatives) depending on the amount and kind of data available and on the size and choice of the representative set.

## 2.4  Learning from Linear Functionals

We can combine the sum-kernel and the derivative kernel naturally and write a single least squares problem for the coefficients in (1) that is solved to fit an interatomic potential to all available total energy, force, and virial stress data (the only condition being that the input data has to be expressible as a linear operation on function values). We define $\mathbf{y}$ as the vector with $L$ components containing all the input data: all total energies, forces, and virial stress components in the training database, $\mathbf{y}'$ as the vector with $N$ components containing the *unknown* atomic energies of the $N$ atomic environments in the database, and $\hat{\mathbf{L}}$ as the linear differential operator of size $L \times N$ which connects $\mathbf{y}$ with $\mathbf{y}'$ such that $\mathbf{y}' = \hat{\mathbf{L}}^{\top}\mathbf{y}$ (note that the definition of $\hat{\mathbf{L}}$ we use here is the transpose of that in Bartók and Csányi 2015). The least squares problem is now to minimize:

$$\|\hat{\mathbf{L}}\mathbf{K}\mathbf{x} - \mathbf{y}\|^2, \tag{22}$$

and the expression for the coefficients is given by:

$$\mathbf{x} = \left[\mathbf{K}_{MM} + (\hat{\mathbf{L}}\mathbf{K}_{NM})^\top \boldsymbol{\Lambda}^{-1} \hat{\mathbf{L}}\mathbf{K}_{NM}\right]^{-1} (\hat{\mathbf{L}}\mathbf{K}_{NM})^\top \boldsymbol{\Lambda}^{-1}\mathbf{y}. \tag{23}$$

It is instructive to write down the above matrices for the simple case when the system consists of just two atoms, $A$ and $B$, with position vectors $\mathbf{r}_A$, $\mathbf{r}_B$, target total energy $E$, and target forces $\mathbf{f}_A \equiv (f_{Ax}, f_{Ay}, f_{Az})$ and $\mathbf{f}_B \equiv (f_{Bx}, f_{By}, f_{Bz})$. The data vector is then given by

$$\mathbf{y} = [E \ f_{Ax} \ f_{Ay} \ f_{Az} \ f_{Bx} \ f_{By} \ f_{Bz}]^\top. \tag{24}$$

The aim of the fit is to determine two unknown atomic energy functions $\varepsilon_A$ and $\varepsilon_B$ as a function of the atomic environments centered around the two atoms, $\mathcal{A}$ and $\mathcal{B}$, respectively. The total energy is their sum, $E = \varepsilon_A + \varepsilon_B$, and the forces need to include the cross terms,

$$
\begin{aligned}
\mathbf{f}_A &= \frac{\partial \varepsilon_A}{\partial \mathbf{r}_A} + \frac{\partial \varepsilon_B}{\partial \mathbf{r}_A}, \\
\mathbf{f}_B &= \frac{\partial \varepsilon_A}{\partial \mathbf{r}_B} + \frac{\partial \varepsilon_B}{\partial \mathbf{r}_B}.
\end{aligned}
\tag{25}
$$

The representative set in this case consists of the same two atoms, so $N = M$, and the kernel matrix is square,

$$\mathbf{K} = \begin{bmatrix} K(\mathcal{A}, \mathcal{A}) & K(\mathcal{A}, \mathcal{B}) \\ K(\mathcal{B}, \mathcal{A}) & K(\mathcal{B}, \mathcal{B}) \end{bmatrix}, \tag{26}$$

and the linear operator $\hat{\mathbf{L}}$ is a $7 \times 2$ matrix and is given by

$$\hat{\mathbf{L}} = \begin{bmatrix} 1 & 1 \\ \overleftarrow{\nabla}_{\mathbf{r}_A} & \overleftarrow{\nabla}_{\mathbf{r}_A} \\ \overleftarrow{\nabla}_{\mathbf{r}_B} & \overleftarrow{\nabla}_{\mathbf{r}_B} \end{bmatrix}, \tag{27}$$

so the $\hat{\mathbf{L}}\mathbf{K}$ matrix to be substituted into Eq. (23) is

$$\hat{\mathbf{L}}\mathbf{K} = \begin{bmatrix} K(\mathcal{A}, \mathcal{A}) + K(\mathcal{A}, \mathcal{B}) & K(\mathcal{B}, \mathcal{A}) + K(\mathcal{B}, \mathcal{B}) \\ \overleftarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{A}, \mathcal{A}) + \overleftarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{B}, \mathcal{A}) & \overleftarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{A}, \mathcal{B}) + \overleftarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{B}, \mathcal{B}) \\ \overleftarrow{\nabla}_{\mathbf{r}_B} K(\mathcal{A}, \mathcal{A}) + \overleftarrow{\nabla}_{\mathbf{r}_B} K(\mathcal{B}, \mathcal{A}) & \overleftarrow{\nabla}_{\mathbf{r}_B} K(\mathcal{A}, \mathcal{B}) + \overleftarrow{\nabla}_{\mathbf{r}_B} K(\mathcal{B}, \mathcal{B}) \end{bmatrix} \tag{28}$$

Note that terms such as $\overleftarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{B}, \mathcal{B})$ or $\overleftarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{A}, \mathcal{B})$ are not zero because atom $A$ is present in the environment $\mathcal{B}$ of atom $B$, and so $K(\mathcal{B}, \mathcal{A})$, and also $K(\mathcal{B}, \mathcal{B})$ depend on $\mathbf{r}_A$ explicitly.

Using the approach of Chmiela et al. (2017) for the dimer, the kernel matrix is $6 \times 6$ and is given by:

$$\mathbf{K}_{\nabla \mathcal{A} \nabla \mathcal{B}} = \begin{bmatrix} \overleftarrow{\nabla}_{\mathbf{r}_A} \overrightarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{A}, \mathcal{A}) & \overleftarrow{\nabla}_{\mathbf{r}_A} \overrightarrow{\nabla}_{\mathbf{r}_B} K(\mathcal{A}, \mathcal{B}) \\ \overleftarrow{\nabla}_{\mathbf{r}_B} \overrightarrow{\nabla}_{\mathbf{r}_A} K(\mathcal{B}, \mathcal{A}) & \overleftarrow{\nabla}_{\mathbf{r}_B} \overrightarrow{\nabla}_{\mathbf{r}_B} K(\mathcal{B}, \mathcal{B}) \end{bmatrix}. \tag{29}$$

In practice it is always worth using *all* available data, even though once the fit is converged in the limit of infinite amount of data, the information from derivatives (forces) is the same as from energies. With finite amount of data, however, choosing the weights corresponding to energies and forces via the diagonal regulariser allows control of the fit, in the sense of its relative accuracy in reproducing energies and forces.

## 3 Density-Based Representations and Kernels

Having summarized the algorithms that can be used to perform kernel ridge regression using atomic-scale properties and their derivatives as inputs, we now proceed to describe a framework for defining physics-based descriptors of local atomic environments and the kernels built from them. In kernel ridge regression, the descriptors do not necessarily need to be expressed explicitly but can also be defined implicitly by means of the kernel function $K(\mathcal{A}, \mathcal{B})$ that corresponds to the scalar product of descriptor vectors that span a (possibly infinite-dimensional) Hilbert space (Schölkopf and Smola 2002). Vectors $|\mathcal{A}\rangle$ in this "reproducing kernel Hilbert space" do correspond to atomic structures, and one can write formally $K(\mathcal{A}, \mathcal{B}) \equiv \langle \mathcal{A}|\mathcal{B} \rangle$ even if the kernel might be computed without ever determining the vectors explicitly.

The reader trained in quantum mechanics will recognize an isomorphism between descriptors and the state vectors on one hand and kernels and expectation values on the other. This analogy suggests that it may be beneficial to formulate atomic-scale descriptors using a formalism that mimics Dirac notation. Whereas in a quantum mechanical setting the physical symmetries of the problem are built into the Hamiltonian, in a machine learning setting, they are more conveniently included in the representation itself that should be made invariant to basic symmetries such as atom labeling, rigid translations, and rotations. In this section we show how starting from these intuitions one can build a very abstract description of a molecular structure that is naturally invariant with respect to the physical symmetries, based on a representation of the atom density.

Translational and rotational symmetries can be included by decomposing the structure into a collection of local environments and by explicit symmetrization over the $SO(3)$ group. This construction is reminiscent of methods that have been applied to the comparison of solid objects (Kazhdan et al. 2003) and leads naturally to the SOAP descriptor and kernel (Bartók et al. 2013b) and to several other popular choices of density-based descriptors – from Behler-Parrinello symmetry functions (Behler and Parrinello 2007) to voxel density descriptors (Kajita et al. 2017) to the binning of the pair correlation function (Schütt et al. 2014) – that

can be regarded as different projections of the same smooth atomic amplitude. A peculiarity of the SOAP framework is that it allows one to explicitly compute the descriptors whose scalar product constitutes the kernel function, which allows one to go back and forth between a kernel and a descriptor language. The atomic environmental descriptors can then be modified to generate nonlinear kernels, as well as combined into global structural kernels. We will briefly discuss different possible approaches to the latter, either by simple linear combination of the local descriptors, or by a more sophisticated procedure that takes into account the most effective matching between pairs of environments in the two structures that are being compared.

## 3.1  A Dirac Notation for Structural Descriptors

Let us introduce an abstract notation to describe atomistic structures in terms of the positions and chemical nature of the atoms that compose them. Taking inspiration from Dirac notation for quantum mechanical states, we associate a ket $|\mathcal{A}\rangle$ with each configuration. Let us start with a simple example to see how such a formalism can be introduced and used. Much like in the case of quantum states, we can define a concrete representation of the ket associated with a structure in terms of positions and chemical species, e.g.:

$$\langle \mathbf{r}|\mathcal{A}\rangle = \sum_i g_i(\mathbf{r} - \mathbf{r}_i)\,|\alpha_i\rangle, \tag{30}$$

where the position of each atom is represented by a smooth density $g_i$ (that in principle could depend on the nuclear charge and the position of atom $i$) and the kets $|\alpha_i\rangle$ contain the information on the nuclear charge of each atom.

The Dirac notation lends itself naturally to the definition of overlap kernels between structures, $\langle \mathcal{A}|\mathcal{B}\rangle$. To compute such an integral, one can use the position representation and assume that the kets associated with different elements are orthonormal:

$$
\begin{aligned}
\langle \mathcal{A}|\mathcal{B}\rangle &= \int d\mathbf{r}\,\langle \mathcal{A}|\mathbf{r}\rangle\,\langle \mathbf{r}|\mathcal{B}\rangle \\
&= \sum_{ij} \int d\mathbf{r}\, g_i^A(\mathbf{r} - \mathbf{r}_i^A)^\star g_j^B(\mathbf{r} - \mathbf{r}_j^B)\left\langle \alpha_i^A \middle| \alpha_j^B \right\rangle \\
&= \sum_\alpha \sum_{i,\,j \in \{\alpha\}} \int d\mathbf{r}\, g_i^A(\mathbf{r} - \mathbf{r}_i^A)^\star g_j^B(\mathbf{r} - \mathbf{r}_j^B).
\end{aligned}
\tag{31}
$$

This density-based representation would not be in itself very useful, as the kernel is not invariant to relative rotations of the structures, and not even to the absolute

position of the two structures in space, or their periodic representation. Nevertheless, it can be taken as the starting point to introduce many of the most successful feature descriptors that have been used in the recent years for machine learning of materials and molecules. To see how, one can take inspiration from linear-scaling electronic structure methods and the nearsightedness principle for electronic matter (Yang 1991; Galli and Parrinello 1992; Goedecker 1999; Prodan and Kohn 2005). We then shift the attention from the description of complete structures to that of spherical atomic environments that one can conveniently center on top of each atom. This is also consistent with the atom-centered potentials that have been discussed in the previous section as an obvious application of this framework. We will use the notation $\left|\mathcal{X}_j\right\rangle$ to indicate an environment centered around the $j$-th atom in a structure and express it in the position representation as:

$$\left\langle \mathbf{r} \middle| \mathcal{X}_j \right\rangle = \sum_i f_c(r_{ij}) g_{ij}(\mathbf{r} - \mathbf{r}_{ij}) \left| \alpha_i \right\rangle \tag{32}$$

where $f_c(r_{ij})$ is a cutoff function that restricts the environment to a spherical region centered on the atom for the sake of computational efficiency and/or localization of the density information. The atom-centered smoothing functions are typically taken to be uniform-width Gaussians, but it would be easy to generalize the expression to include a dependency on the atomic species and/or the distance of an atom from the center of the environment, which could be used to, e.g., reduce the resolution of the descriptor at the periphery of the environment or adapt the smoothing length scale to each atomic species.

Note that one could also combine the density contributions from atoms of the same species into a species-dependent atomic amplitude,

$$\left\langle \alpha \mathbf{r} \middle| \mathcal{X}_j \right\rangle = \psi_{\mathcal{X}_j}^\alpha(\mathbf{r}) = \sum_{i \in \alpha} f_c(r_{ij}) g_{ij}(\mathbf{r} - \mathbf{r}_{ij}), \tag{33}$$

and then write

$$\left\langle \mathbf{r} \middle| \mathcal{X}_j \right\rangle = \sum_\alpha \psi_{\mathcal{X}_j}^\alpha(\mathbf{r}) \left| \alpha \right\rangle. \tag{34}$$

This notation is very useful to reveal how different descriptors can be seen as alternative representations of the same abstract ket. For instance, one can expand the atom density in orthogonal radial functions $R_n(r)$ and spherical harmonics. The coefficients in such an expansion can be written as

$$\left\langle \alpha n l m \middle| \mathcal{X}_j \right\rangle = \int d\mathbf{r} \left\langle n l m | \mathbf{r} \right\rangle \left\langle \alpha \mathbf{r} \middle| \mathcal{X}_j \right\rangle$$

$$= \int dr \, d\hat{\mathbf{r}} \, r^2 R_n(r) Y_m^l(\hat{\mathbf{r}}) \psi_{\mathcal{X}_j}^\alpha(r\hat{\mathbf{r}}). \tag{35}$$

As another example, Behler-Parrinello atom-centered symmetry functions that have been used in the construction of artificial neural network-based interatomic potentials for materials (Behler and Parrinello 2007; Eshet et al. 2012; Morawietz et al. 2016; Cheng et al. 2016) and molecules (Smith et al. 2017) can be written by setting the basis functions to be delta distributions $g_{ij}(\mathbf{r} - \mathbf{r}_{ij}) = \delta(\mathbf{r} - \mathbf{r}_{ij})$ and averaging the atom density with an appropriate pair weighting function $G_2$, e.g.:

$$
\begin{aligned}
\langle \alpha \beta G_2 | \mathcal{X}_j \rangle &= \langle \alpha | \alpha_j \rangle \int d\mathbf{r}\, G_2(r) \langle \beta \mathbf{r} | \mathcal{X}_j \rangle \\
&= \delta_{\alpha_j \alpha} \sum_{i \in \{\beta\}} f_c(r_{ij}) G_2(r_{ij})
\end{aligned}
\tag{36}
$$

The basis functions of the spectral neighbor analysis potential (Thompson et al. 2015) also start with the same density and expands it in hyperspherical harmonics as introduced in Bartók et al. (2010).

## 3.2   Smooth Overlap of Atomic Positions

It is clear that a density-based representation such as Eq. (32) is invariant to translations of the entire structure, but not to rotations that would change the orientation of the atomic neighbor amplitude. This reflects the fact that scalar products of the form $\langle \mathcal{X}_j | \mathcal{X}_k \rangle$ depend on the relative orientation of the environments being compared. In the smooth overlap of atomic positions (SOAP) framework, we define a symmetrized version of the overlap kernel, using the Haar integral (Haar 1933) of the rotation group:

$$
K^{(\nu)}(\mathcal{X}_j, \mathcal{X}_k) = \int d\hat{R}\, |\langle \mathcal{X}_j | R \rangle | \mathcal{X}_k \rangle|^{\nu} = \left\langle \mathcal{X}_j^{(\nu)} \middle| \mathcal{X}_k^{(\nu)} \right\rangle
\tag{37}
$$

where the integral is performed over all possible rotation matrices. If the base kernel is raised to the $\nu$-th power, the average preserves information on the correlations between atoms up to the $(\nu + 1)$-th order (Glielmo et al. 2018). As we will show below, a crucial feature of the SOAP framework is that an explicit expression for the symmetrized descriptor vectors $|\mathcal{X}_j^{(\nu)}\rangle$ can be given, which is quite manageable for $\nu = 1, 2$, but becomes increasingly cumbersome for higher $\nu$. An effective description of higher-order interactions that does not increase too much the complexity of the analytical evaluation of (37) can be obtained by manipulating the $\nu = 2$ kernel, e.g., by taking a nonlinear function of it. In practice it has been found that raising it to a power $\zeta$ and normalizing it to 1:

$$
\left\langle \mathcal{X}_j^{(2)} \middle| \mathcal{X}_k^{(2)} \right\rangle_{\zeta} = \frac{\left\langle \mathcal{X}_j^{(2)} | \mathcal{X}_k^{(2)} \right\rangle^{\zeta}}{\sqrt{\left\langle \mathcal{X}_j^{(2)} | \mathcal{X}_j^{(2)} \right\rangle^{\zeta} \left\langle \mathcal{X}_k^{(2)} | \mathcal{X}_k^{(2)} \right\rangle^{\zeta}}}
\tag{38}
$$

are sufficient to include many-body contributions in the final kernel.

Using the Dirac notation, it is easy to see how one can give an explicit representation of the $SO(3)$ symmetrized ket for the case with $\nu = 1, 2$. Using a spherical harmonics expansion of $|\mathcal{X}_j\rangle$, it is very natural to perform the rotational average analytically by introducing the Wigner matrix associated with the rotation: $\langle lm| \hat{R} |l'm'\rangle = \delta_{ll'} D^l_{mm'}(\hat{R})$

$$\int \mathrm{d}\hat{R} \sum_{\alpha nlm} \langle \mathcal{X}_j|\alpha nlm\rangle \langle \alpha nlm| \hat{R} |\mathcal{X}_k\rangle = \sum_{\alpha nlmm'} \langle \mathcal{X}_j|\alpha nlm\rangle \langle \alpha nlm'|\mathcal{X}_k\rangle \int \mathrm{d}\hat{R}\, D^l_{mm'}(\hat{R}) \tag{39}$$

which simplifies greatly due to the properties of the Wigner matrices. Only the term with $l = 0$ survives, which makes it possible to write explicitly the $\nu = 1$ symmetrized SOAP descriptors in terms of the spherical harmonics coefficients

$$\left\langle \alpha n \middle| \mathcal{X}_k^{(1)} \right\rangle = \sqrt{8\pi^2}\, \langle \alpha n00|\mathcal{X}_k\rangle, \tag{40}$$

which corresponds to the simple kernel

$$\left\langle \mathcal{X}_j^{(1)} \middle| \mathcal{X}_k^{(1)} \right\rangle = \sum_{\alpha n} \left\langle \mathcal{X}_j^{(1)} \middle| \alpha n \right\rangle \left\langle \alpha n \middle| \mathcal{X}_k^{(1)} \right\rangle. \tag{41}$$

A position representation of the $\nu = 1$ descriptor $\langle r|\mathcal{X}_k^{(1)}\rangle$ yields naturally the rotational average of $\langle \mathbf{r}|\mathcal{X}_k\rangle$. This can be seen by expressing $K^{(1)}(\mathcal{X}_j, \mathcal{X}_k)$ in a position basis

$$\left\langle \alpha \mathcal{X}_j^{(1)} \middle| \alpha \mathcal{X}_k^{(1)} \right\rangle = \int \mathrm{d}\hat{R} \int \mathrm{d}\mathbf{r}\, \psi^\alpha_{\mathcal{X}_j}(\mathbf{r}) \psi^\alpha_{\mathcal{X}_k}(\hat{R}\mathbf{r})$$
$$= 32\pi^3 \int \mathrm{d}r\, r^2 \bar{\psi}^\alpha_{\mathcal{X}_j}(r) \bar{\psi}^\alpha_{\mathcal{X}_k}(r) \tag{42}$$

where we have defined the rotationally average atom density:

$$\bar{\psi}^\alpha_{\mathcal{X}_j}(r) = \frac{1}{4\pi} \int \mathrm{d}\hat{\mathbf{r}}\, \psi^\alpha_{\mathcal{X}_j}(r\hat{\mathbf{r}}) = \frac{1}{\sqrt{32\pi r^3}} \left\langle \alpha r \middle| \mathcal{X}_j^{(1)} \right\rangle, \tag{43}$$

which is thus closely related to the pair correlation function around the tagged atom. Similar descriptors have been used for machine learning of molecules and materials (Schütt et al. 2014; Faber et al. 2015), revealing once more the intimate relationships between different atom density-based descriptors.

The $\nu = 1$ descriptor integrates away all angular correlations and therefore does not provide a unique representation of an environment. The descriptors with $\nu = 2$ provide information on three-body correlations and can also be obtained relatively

easily in closed form. The Haar integral now contains the product of two Wigner matrices. Exploiting their orthogonality relations, one obtains

$$
\int d\hat{R} \left| \sum_{\alpha n l m} \langle \mathcal{X}_j | \alpha n l m \rangle \langle \alpha n l m | \hat{R} | \mathcal{X}_k \rangle \right|^2 = \sum_{\alpha n \alpha' n' l} \left\langle \mathcal{X}_j^{(2)} \middle| \alpha n \alpha' n' l \right\rangle \left\langle \alpha n \alpha' n' l \middle| \mathcal{X}_k^{(2)} \right\rangle
$$

(44)

where the $\nu = 2$ symmetrized SOAP descriptors read

$$
\left\langle \alpha n \alpha' n' l \middle| \mathcal{X}_j^{(2)} \right\rangle = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m \langle \mathcal{X}_j | \alpha n l m \rangle \langle \alpha' n' l m | \mathcal{X}_j \rangle .
$$

(45)

This notation corresponds to the power spectrum components introduced in Bartók et al. (2013b) and De et al. (2016), $\left\langle \alpha n \alpha' n' l \middle| \mathcal{X}_j^{(2)} \right\rangle \equiv p_{nn'l}^{\alpha\alpha'}(\mathcal{X}_j)$. Note also that, while the representation of the symmetrized kets in terms of the $nlm$ expansion is very convenient, it is not the only possibility. Similar to Eq. (43), an explicit position representation can be obtained for $\left\langle \alpha \mathbf{r}_1 \alpha' \mathbf{r}_2 \middle| \mathcal{X}_k^{(2)} \right\rangle$ that provides a complete representation of the three-body rotationally invariant correlations. The three-body symmetry functions of the Behler-Parrinello kind can be seen as projections of this representation, similar to the case of two-body functions in Eq. (36).

The case of $\nu = 3$ leads to an explicit representation of the ket that is proportional to the bispectrum of the environment (Bartók et al. 2013b)

$$
\left\langle \alpha_1 n_1 l_1 \alpha_2 n_2 l_2 \alpha n l \middle| \mathcal{X}_j^{(3)} \right\rangle \propto \sum_{m m_1 m_2} \langle \mathcal{X}_j | \alpha n l m \rangle
$$

$$
\times \left\langle \alpha_1 n_1 l_1 m_1 \middle| \mathcal{X}_j \right\rangle \left\langle \alpha_2 n_2 l_2 m_2 \middle| \mathcal{X}_j \right\rangle \langle l_1 \, m_1 \, l_2 \, m_2 | l \, m \rangle .
$$

(46)

While the dimensionality of this descriptor makes it impractical unless somehow sparsified, it does give direct access to higher-order correlations. An interesting detail is that $\left| \mathcal{X}_j^{(3)} \right\rangle$, contrary to the $\nu = 1, 2$ cases, is not invariant to mirror symmetry, which makes it capable of distinguishing enantiomers.

Finally, one should note that the normalization of the kernel Eq. (38) can be achieved by normalizing the SOAP vector, so that an explicit representation of the normalized descriptor is possible. While in principle one could write out an explicit representation that yields the kernel for $\zeta > 1$, it would contain an exponentially increasing number of terms. As in the case of $\left| \mathcal{X}_j^{(3)} \right\rangle$, this only makes sense if combined with a sparsification procedure.

## 3.3   λ-SOAP: Symmetry-Adapted Gaussian Process Regression

When building a machine learning model for a tensorial property **T**, one should consider that the target is not invariant under the action of a symmetry operation (e.g., a rotation) but transforms covariantly. The most effective strategy to encode

the appropriate covariance properties in the model involves the decomposition of the tensor into its irreducible spherical components, i.e., combinations of the elements of the tensor that transform as the spherical harmonics of order $\lambda$ (Varshalovich et al. 1988). For these irreducible components,

$$T_{\lambda\mu}(\hat{R}\mathcal{X}_j) = \sum_{\mu'} D^\lambda_{\mu\mu'}(\hat{R}) T_{\lambda\mu'}(\mathcal{X}_j) \tag{47}$$

As shown in Glielmo et al. (2017) for the case of vectors and in Grisafi et al. (2018) for tensors of arbitrary order, one has to consider a matrix-valued kernel that describes the geometric relationship between the different components of $\mathbf{T}_\lambda$, which can be obtained by including an additional Wigner matrix $D^\lambda_{\mu\mu'}(\hat{R})$ in the Haar integral

$$\left\langle \mathcal{X}^{(v)}_{j,\lambda\mu} \middle| \mathcal{X}^{(v)}_{k,\lambda\mu'} \right\rangle = \int d\hat{R} \, D^\lambda_{\mu\mu'}(\hat{R}) \left| \langle \mathcal{X}_j | R \rangle | \mathcal{X}_k \rangle \right|^v. \tag{48}$$

For the case with $v = 2$, the symmetrized kets can be written explicitly based on a $\alpha nlm$ expansion of the atom density:

$$\left\langle \alpha n l \alpha' n' l' \middle| \mathcal{X}^{(2)}_{j,\lambda\mu} \right\rangle = \sqrt{\frac{8\pi^2}{2l+1}} \sum_{mm'} \langle \mathcal{X}_j | \alpha n l m \rangle$$
$$\times \langle \alpha' n' l' m' | \mathcal{X}_j \rangle \langle l \, m \, l' \, -m' | \lambda \, -\mu \rangle \tag{49}$$

We write Eq. (49) in this form because it is somewhat symmetric, but the properties of the CG coefficients require that $m' = m + \mu$, so the expression can be evaluated with a single sum. Furthermore, the expression evaluates to 0 whenever $|l - l'| < \lambda$, which reduces the number of elements that must be evaluated and stored and makes it clear that Eq. (49) reduces to the scalar SOAP Eq. (45) when $\lambda = 0$.

When using a linear model, each of the symmetry-adapted descriptors Eq. (49) can be used to represent tensorial components that transform as $Y^\lambda_\mu$. Linearity, in this case, is necessary for preserving the symmetry properties of the $\lambda$-SOAP. A nonlinear model, however, can be obtained by scaling each $\left\langle \alpha n l \alpha' n' l' \middle| \mathcal{X}^{(2)}_{j,\lambda\mu} \right\rangle$ by a (in principle different) nonlinear function of some $\lambda = 0$ descriptors. In the kernel language, a high-order version of the $\lambda$-SOAP kernel can be introduced with an expression analogous to (38):

$$\left\langle \mathcal{X}^{(2)}_{j,\lambda\mu} \middle| \mathcal{X}^{(2)}_{k,\lambda\mu'} \right\rangle_\zeta = \frac{\left\langle \mathcal{X}^{(2)}_{j,\lambda\mu} \middle| \mathcal{X}^{(2)}_{k,\lambda\mu'} \right\rangle \langle \mathcal{X}_j^{(2)} | \mathcal{X}_k^{(2)} \rangle_{\zeta-1}}{\left\| \left\langle \mathcal{X}^{(2)}_{j,\lambda\boldsymbol{\mu}} \middle| \mathcal{X}^{(2)}_{j,\lambda\boldsymbol{\mu}^\top} \right\rangle \right\|_F \left\| \left\langle \mathcal{X}^{(2)}_{k,\lambda\boldsymbol{\mu}} \middle| \mathcal{X}^{(2)}_{k,\lambda\boldsymbol{\mu}^\top} \right\rangle \right\|_F}, \tag{50}$$

where $\|\cdot\|_F$ indicates the Frobenius norm and $\left\langle \mathcal{X}_j^{(2)} \middle| \mathcal{X}_k^{(2)} \right\rangle_{\zeta-1}$ is a (scalar) SOAP kernel. This second term makes the overall kernel nonlinear, without affecting the symmetry properties of the overall tensorial kernel.

## 3.4   Kernel Operators and Descriptor Optimization

Provided one takes a long-range environmental cutoff and chooses a kernel that can represent high orders of many-body interactions, a density-based representation of atomic structures should provide a complete description of any atomic structure and – given a sufficiently complete training set – predict any atomistic property with arbitrary accuracy. In practice, obviously, the accuracy of a model depends on the details of the representation, which is why different descriptors or kernels provide different levels of accuracy for the same training and test set (Faber et al. 2017). The performance of a set of descriptors can be improved by modifying them so that they represent more efficiently the relations between structure and properties.

The relation between structures, properties, and kernels can be encoded in a Hermitian operator $\hat{W}$ bracketed between the translationally invariant environmental kets:

$$\langle \mathcal{X}_j | \mathcal{X}_k \rangle_{\hat{W}} \equiv \langle \mathcal{X}_j | W \rangle | \mathcal{X}_k \rangle. \tag{51}$$

When the kernel is averaged over rotations to obtain the SOAP kernel and the associated descriptors,

$$\left\langle \mathcal{X}_j^{(v)} \middle| \mathcal{X}_k^{(v)} \right\rangle_{\hat{W}} = \int d\hat{R} \left| \langle \mathcal{X}_j | W \rangle \hat{R} | \mathcal{X}_k \rangle \right|^p \tag{52}$$

it becomes clear that if we want to be able to write the "operator-adapted" $SO(3)$ descriptors explicitly, the operator $\hat{W}$ must commute with rotations. This additional requirement means that the most general form of $\hat{W}$, written in the $\{|\alpha nlm\rangle\}$ basis, is

$$\langle \alpha nlm | W \rangle | \alpha'n'l'm' \rangle = \delta_{ll'}\delta_{mm'}\hat{W}_{\alpha\alpha'nn'll'}. \tag{53}$$

This operator can be used, for instance, to scale the density as a function of distance or to introduce an "alchemical kernel" that couples the descriptors associated with different elements and in principle to realize an even more subtle tuning of the SOAP descriptors, in which radial, angular, and alchemical channels are mixed to provide a more efficient representation of complex chemical environments.

## 3.5  Computing SOAP Descriptors Efficiently

A practical calculation of both scalar and tensorial $\nu = 2$ SOAP descriptors $\left\langle \alpha n l \alpha' n' l' \middle| \mathcal{X}_{j,\lambda\mu}^{(2)} \right\rangle$ requires the evaluation of the expansion coefficients $\langle \alpha n l m | \mathcal{X}_j \rangle$. Let us start with the atom density written in the position representation, according to Eq. (35), and consider the case in which $\psi_{\mathcal{X}}^{\alpha}(\mathbf{r})$ is written as a superposition of spherical Gaussian functions of width $\sigma$ placed at the positions of the atoms of type $\alpha$. Then, the spherical harmonic projection in Eq. (35) can be carried out analytically, leading to:

$$
\langle \alpha n l m | \mathcal{X}_j \rangle = \sum_{i \in \alpha} Y_{lm}(\hat{\mathbf{r}}_{ij})\, e^{-\frac{r_{ij}^2}{2\sigma^2}} \times
$$
$$
\times \int_0^\infty \mathrm{d}r \; r^2 \; R_n(r) e^{-\frac{r^2}{2\sigma^2}} \iota_l\left(\frac{r r_{ij}}{\sigma^2}\right) \tag{54}
$$

where the sum runs over all neighboring atoms of type $\alpha$ and $\iota_l$ indicates a modified spherical Bessel function of the first kind. It is convenient to choose a form for the orthogonal radial basis functions $R_n(r)$ that makes it possible to perform the radial integration analytically.

One possible choice starts by using Gaussian type orbitals as non-orthogonal primitive functions $\tilde{R}_k(r)$

$$
\tilde{R}_k(r) = \mathcal{N}_k \, r^k \exp\left\{ -\frac{1}{2}\left(\frac{r}{\sigma_k}\right)^2 \right\}, \tag{55}
$$

where $\mathcal{N}_k$ is a normalization factor, such that $\int_0^\infty dr\, r^2 \tilde{R}_k^2(r) = 1$. The set of Gaussian widths $\{\sigma_k\}$ can be chosen to span effectively the radial interval involved in the environment definition. Assuming that the smooth cutoff function approaches one at a distance $r_{\text{cut}} - \delta r_{\text{cut}}$, one could take $\sigma_k = (r_{\text{cut}} - \delta r_{\text{cut}}) \max(\sqrt{k}, 1)/n_{\max}$ that gives functions that are peaked at equally spaced positions in the range between 0 and $r_{\text{cut}} - \delta r_{\text{cut}}$.

While the $\tilde{R}_k(r)$ are not themselves orthogonal, they can be used to write orthogonal basis functions $R_n(r) = \sum_k S_{nk}^{-1/2} \tilde{R}_k(r)$, where the overlap matrix $S_{kk'} = \int dr\, r^2 \tilde{R}_k(r) \tilde{R}_{k'}(r)$ can be computed analytically. The full decomposition of the translationally invariant environmental ket can then be obtained without recourse to numerical integration.

Once the spherical decomposition of the atomic density has been obtained, the coefficients can be combined to give the SOAP descriptors of orders 1 and 2. Particularly in the presence of many different chemical species, the number of components can become enormous. Ignoring for simplicity a few symmetries, and the fact that if all species do not appear in every environment, it is possible

to store a sparse representation of the descriptor, the power spectrum contains a number of components of the order of $n_{\text{species}}^2 n_{\text{max}}^2 l_{\text{max}}$, which can easily reach into the tens of thousands. In the case of the tensorial $\lambda$-SOAP, the number increases further to $\lambda^2 n_{\text{species}}^2 n_{\text{max}}^2 l_{\text{max}}$. It is however not necessary to compute and store all of these descriptors: each of them, or any linear combination, is a spherical invariant (covariant) description of the environment and can be used separately as a descriptor. This can be exploited to reduce dramatically the computational cost and the memory footprint of a SOAP calculation, determining a low-rank approximation of the descriptor. One can use dimensionality reduction techniques similar to those discussed in Sect. 2.1 to identify the most suitable reference structures. As shown in Imbalzano et al. (2018), both CUR decomposition and a greedy selection strategy based on farthest point sampling make it possible to reduce by more than 95% the number of SOAP descriptors that are needed to predict the energy of small organic molecules with chemical accuracy.

## 3.6    Back to the Structures

Whenever one is interested in computing properties that are associated to individual atoms (for instance, their NMR chemical shieldings or the forces), one can use directly the descriptors corresponding to each environment or the kernel between two environments, as the basis for a linear or nonlinear regression model. As discussed in Sect. 2, it is often the case that one is interested in using as structure labels some properties that are instead associated with the entirety of a structure, e.g., its cohesive energy, its dielectric constant, etc. In these cases a ridge regression model should be used that is based on "global" kernels between the structures, $K(\mathcal{A}, \mathcal{B})$, rather than those between individual atom-centered environments. This is reflected in how the kernels between environments should be combined to give a kernel that is suitable to represent the relation between local environments and the overall property of a structure. When the target property can be seen as an additive combination of local, atom-centered contributions, the most natural (and straightforward) choice, that is consistent with Eq. (13), is

$$K(\mathcal{A}, \mathcal{B}) = \sum_{j \in \mathcal{A}, k \in \mathcal{B}} K(\mathcal{X}_j, \mathcal{X}_k). \tag{56}$$

It is worth stressing that in the case where the environment kernel is a linear kernel based on SOAP descriptors, this sum-kernel can be written in terms of a global descriptor associated with the entire structure,

$$K(\mathcal{A}, \mathcal{B}) = \left\langle \mathcal{A}^{(\nu)} \middle| \mathcal{B}^{(\nu)} \right\rangle, \tag{57}$$

where we introduced

$$\left|\mathcal{A}^{(\nu)}\right\rangle = \sum_{j \in \mathcal{A}} \left|\mathcal{X}_j^{(\nu)}\right\rangle. \tag{58}$$

An alternative way to combine the information from individual environments in a symmetrized global kernel corresponds to averaging the Fourier coefficients of each environment,

$$\langle \alpha nlm | \mathcal{A} \rangle = \sum_{j \in \mathcal{A}} \langle \alpha nlm | \mathcal{X}_j \rangle \tag{59}$$

and then taking the Haar integral of the resulting sum. For instance, for $\nu = 2$,

$$\left\langle \alpha n\alpha'n'l \middle| \bar{\mathcal{A}}^{(2)} \right\rangle = \sum_m \langle \alpha nlm | \mathcal{A} \rangle \langle \mathcal{A} | \alpha'n'lm \rangle. \tag{60}$$

The form Eq. (56) is more general, and one can readily introduce nonlinear kernels such as $\left\langle \mathcal{X}_j^{(\nu)} \middle| \mathcal{X}_k^{(\nu)} \right\rangle_\zeta$ for which an explicit expression for the descriptors would be too cumbersome. Equation (56) also suggests that the combination of environment kernels could be generalized by introducing a weighting matrix

$$K_W(\mathcal{A}, \mathcal{B}) = \sum_{j \in \mathcal{A}, k \in \mathcal{B}} W_{jk}(\mathcal{A}, \mathcal{B}) K(\mathcal{X}_j, \mathcal{X}_k). \tag{61}$$

One could, for instance, determine the importance of each environment within a structure and set $W_{jk}(\mathcal{A}, \mathcal{B}) = w_j(\mathcal{A}) w_k(\mathcal{B})$. Alternatively, one can use techniques from optimal transport theory (Cuturi 2013) to define an entropy-regularized matching (REMatch) procedure (De et al. 2016), in which $W_{jk}$ is a doubly stochastic matrix that matches the most similar environments in the two structures, disregarding the environmental kernels between very dissimilar environments

$$\mathbf{W}(\mathcal{A}, \mathcal{B}) = \operatorname*{argmin}_{\mathbf{W} \in \mathcal{U}(N_\mathcal{A}, N_\mathcal{B})} \sum_{jk} W_{jk} \left[ d^2(\mathcal{X}_j, \mathcal{X}_k) + \gamma \ln W_{jk} \right], \tag{62}$$

where $d^2$ indicates the kernel-induced squared distance Eq. (9). The parameter $\gamma$ weights the entropy regularization and makes it possible to interpolate between strict matching of the most similar pairs of environments ($\gamma \to 0$) to an average kernel that weights all pairs equally ($\gamma \to \infty$). Although this construction complicates considerably the combination of local kernels, it provides a strategy to introduce an element of non-locality in the comparison between structures. Given the cost of computing the REMatch kernel and the fact that it prevents using some sparsification strategies that act at the level of individual environments, this method should be used when the target property is expected to exhibit very strong nonadditive behavior,

e.g., when just one portion of the system is involved, for instance, when determining the activity of a drug molecule, a problem for which REMatch has been shown to improve dramatically the accuracy of the ML model (Bartók et al. 2017).

## 3.7    Multi-kernel Learning

We have shown that SOAP descriptors can be seen as just one possible embodiment of a general class of rotationally symmetrized density-based descriptors that also encompasses other popular descriptors for atomic-scale machine learning, and that can be tuned to a great extent, e.g., by changing the way different components are weighted. The fact that different descriptors can be computed within the same formalism does not imply they are fully equivalent: each expression or kernel emphasizes different components of the structure/property relations. For instance, kernels with varying radial scaling or cutoff distance focus the machine learning model on short-, mid-, or long-range interactions. It is then natural to consider whether a better overall model can be constructed by combining descriptors that are associated with different cutoff distances or different levels of body order expansions. This can be achieved by a weighted combination of kernels of the form

$$K_{\text{tot}}(\mathcal{A}, \mathcal{B}) = \sum_{\aleph} w_{\aleph} K_{\aleph}(\mathcal{A}, \mathcal{B}), \tag{63}$$

where each $K_{\aleph}$ corresponds to a distinct model.

This is equivalent to an additive model for a property, similar to the construction of an atom-centered decomposition of the total energy in Eq. (12). In this case, instead, the property $y$ associated with each structure is written as the sum of contributions $y_{\aleph}(\mathcal{A})$ that are associated with the various kernels $K_{\aleph}$

$$y(\mathcal{A}) = \sum_{\aleph} y_{\aleph}(\mathcal{A}) = \sum_{\aleph, \mathcal{B}} x_{\mathcal{B}} w_{\aleph} K_{\aleph}(\mathcal{A}, \mathcal{B}) \tag{64}$$

where $x_{\mathcal{B}}$ are the kernel regression weights for each of the representative structures $\mathcal{B}$. The weights $w_{\aleph}$ correspond to the estimated contribution that each model will give to the final property and can be obtained by cross-validation or by physical intuition. For instance, in the case of multiple radial cutoffs, it is found that much smaller weights should be associated with long-range kernels, consistent with the fact that distant interactions contribute a small (although often physically relevant) contribution to the total energy (Bartók et al. 2017). It should also be noted that, provided that the descriptors corresponding to the kernels are linearly independent, Eq. (63) effectively corresponds to a feature space of increased dimensionality, obtained by concatenating the descriptors that are – implicitly or explicitly – associated with each kernel.

# 4    Conclusions

We have laid out a mathematical framework, based on the concept of the atomic density, for building descriptors of atomic environments that preserve the geometric symmetries and chemically sensible limits. Coupled with kernel regression, this allows the fitting of complex models of physical properties on the atomic scale, both scalars like interatomic potentials (force fields) and tensors such as multipole moments and quantum mechanical operators. We discuss in general terms how kernel regression can be extended to include a sparse selection of reference structures and to predict and learn from linear functionals of the target property. To leverage the many formal similarities between kernel regression and quantum mechanics, we use a Dirac bra-ket notation to formulate the main results concerning the SOAP descriptors. This notation also helps in making apparent the relationship between SOAP descriptors and other popular density-based approaches to represent atomic structures. The framework can be extended and tuned in many different ways to incorporate insight about the relations between properties, structures, and descriptors. With physical principles such as symmetry and nearsightedness of interactions at its core, we believe this formulation is ideally suited to provide a unified framework to machine learn atomic-scale properties.

# Nomenclature

| | |
|---|---|
| $\mathcal{A}$ | An item – structure, or atomic environment for which one wants to predict a property |
| $K(\mathcal{A}, \mathcal{B})$ | The kernel function computed between items $\mathcal{A}$ and $\mathcal{B}$ |
| $N$ | Number of input structures in the training set |
| $M$ | Number of structures in the representative set |
| $\mathbf{x}$ | The vector of KRR weights, also written as $\mathbf{x}_M$; the weight associated with a structure $\mathcal{B}$ is indicated as $x_{\mathcal{B}}$ |
| $\mathbf{y}$ | The vector containing the values of the target property, also written as $\mathbf{y}_N$. $y_{\mathcal{B}}$ indicates the value for the item $\mathcal{B}$ |
| $\mathbf{K}$ | Kernel matrix |
| $\mathbf{K}_{MN}$ | Slice of the kernel matrix $\mathbf{K}$, corresponding to rows in set $M$ and columns in set $N$ |
| $^{\Sigma}\mathbf{K}$ | Sum-kernel, defined as the sum of the regular kernel over a set of configurations |
| $\overleftarrow{\nabla}$ | Derivative operator applying to the first argument of the kernel matrix |
| $\overrightarrow{\nabla}$ | Derivative operator applying to the second argument of the kernel matrix |
| $\mathbf{K}_{\nabla_{\mathcal{A}}\mathcal{B}}$ | Derivative of the kernel matrix, applying to its first argument, with respect to the coordinates of atoms in structure $\mathcal{A}$, with structure $\mathcal{B}$ as its second argument |

| | |
|---|---|
| $\hat{\mathbf{L}}$ | Linear operator connecting the observed values $\mathbf{y}$ with the unobserved atomic energies $\mathbf{y}'$ |
| $\|\mathcal{A}\rangle$ | An abstract vector that describes the input $\mathcal{A}$ |
| $\langle\mathcal{A}\|\mathcal{B}\rangle$ | The scalar product between the features associated with $\mathcal{A}$ and $\mathcal{B}$. Could be either an explicit scalar product, or an abstract notation equivalent to $K(\mathcal{A}, \mathcal{B})$ |
| $g_i(\mathbf{r})$ | A smooth function – typically a Gaussian that is used to represent the density associated with atom $i$ |
| $\|\alpha\rangle$ | An abstract vector that represents the chemical species $\alpha$ |
| $\mathbf{r}$ | Position in 3D Cartesian coordinates. |
| $r$ | The modulus of the vector $\mathbf{r}$ |
| $\hat{\mathbf{r}}$ | The unit vector $\mathbf{r}/r$ |
| $\mathbf{r}_i$ | Position of the $i$-th atom. |
| $\mathbf{r}_{ij}$ | Displacement vector $\mathbf{r}_i - \mathbf{r}_j$ between the $i$-th and $j$-th atoms |
| $\psi^{\alpha}_{\mathcal{X}_j}(\mathbf{r})$ | The atom density of species $\alpha$ centered around the $j$-th atom |
| $Y^l_m(\hat{\mathbf{r}})$ | The $l, m$-th spherical harmonic |
| $R_n(r)$ | The $n$-th orthogonal radial basis function |
| $\left\|\mathcal{X}_j^{(\nu)}\right\rangle$ | The spherically-averaged SOAP descriptor of order $\nu$ |
| $\left\langle\mathcal{X}_j^{(\nu)}\middle\|\mathcal{X}_k^{(\nu)}\right\rangle_{\zeta}$ | The normalized SOAP kernel of order $\nu$ and nonlinear exponent $\zeta$. |
| $D^l_{mm'}(\hat{R})$ | The Wigner rotation matrix associated with the rotation $\hat{R}$ |
| $\left\langle\alpha n\alpha'n'l\middle\|\mathcal{X}_j^{(2)}\right\rangle$ | The radial/spherical representation of the SOAP $\nu = 2$ vector, corresponding to the power spectrum between species $\alpha$ and $\alpha'$ |
| $\langle l_1\, m_1\, l_2\, m_2\|l\, m\rangle$ | A Clebsch-Gordan coefficient |
| $T_{\lambda\mu}$ | The $\mu$-th component of the irreducible spherical component of order $\lambda$ for the tensorial quantity $\mathbf{T}$ |
| $\left\|\mathcal{X}_{j,\lambda\mu}^{(\nu)}\right\rangle$ | The $\lambda$-SOAP descriptor of order $\nu$, corresponding to the irreducible spherical component $\lambda\mu$ centered on atom $j$ |
| $\left\langle\mathcal{X}_j\middle\|\mathcal{X}_k\right\rangle_{\hat{W}}$ | A kernel transformed by the Hermitian operator $\hat{W}$ |

# References

Bartók AP, Csányi G (2015) Gaussian approximation potentials: A brief tutorial introduction. Int J Quant Chem 116:1051

Bartók AP, Payne MC, Kondor R, Csányi G (2010) Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett 104:136403

Bartók AP, Gillan MJ, Manby FR, Csányi G (2013a) Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water. Phys Rev B 88:054104

Bartók AP, Kondor R, Csányi G (2013b) On representing chemical environments. Phys Rev B 87:184115

Bartók AP, De S, Poelking C, Bernstein N, Kermode JR, Csányi G, Ceriotti M (2017) Machine learning unifies the modeling of materials and molecules. Sci Adv 3:e1701816

Behler J, Parrinello M (2007) Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. Phys Rev Lett 98:146401

Bernstein N, Kermode JR, Csányi G (2009) Hybrid atomistic simulation methods for materials systems. Rep Prog Phys 72:026501

Bishop CM (2016) Pattern recognition and machine learning. Springer, New York

Braams BJ, Bowman JM (2009) Permutationally invariant potential energy surfaces in high dimensionality. Int Rev Phys Chem 28:577–606

Brenner DW (2000) The Art and Science of an Analytic Potential. Phys Status Solidi B 217:23

Caro MA, Deringer VL, Koskinen J, Laurila T, Csányi G (2018) Growth Mechanism and Origin of High $sp^3$ Content in Tetrahedral Amorphous Carbon. Phys Rev Lett 120:166101

Ceriotti M, Tribello GA, Parrinello M (2013) Demonstrating the Transferability and the Descriptive Power of Sketch-Map. J Chem Theory Comput 9:1521

Cheng B, Behler J, Ceriotti M (2016) Nuclear Quantum Effects in Water at the Triple Point: Using Theory as a Link Between Experiments. J Phys Chem Lett 7:2210

Chmiela S, Tkatchenko A, Sauceda HE, Poltavsky I, Schütt KT, Müller K-R (2017) Machine learning of accurate energy-conserving molecular force fields. Sci Adv 3:e1603015

Cuturi M (2013) Sinkhorn distances: lightspeed computation of optical transport. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ (eds) Advances in neural information processing systems 26. Curran Associates, Inc., pp 2292–2300

De S, Bartók AP, Csányi G, Ceriotti M (2016) Comparing molecules and solids across structural and alchemical space. Phys Chem Chem Phys 18:13754

Deringer VL, Csányi G (2017) Machine learning based interatomic potential for amorphous carbon. Phys Rev B 95:094203

Deringer VL, Pickard CJ, Csányi G (2018) Data-Driven Learning of Total and Local Energies in Elemental Boron. Phys Rev Lett 120:156001

Dragoni D, Daff TD, Csányi G, Marzari N (2018) Achieving DFT accuracy with a machine-learning interatomic potential: Thermomechanics and defects in bcc ferromagnetic iron. Phys Rev Mater 2:013808

Eshet H, Khaliullin RZ, Kühne TD, Behler J, Parrinello M (2012) Microscopic origins of the anomalous melting behavior of sodium under high pressure. Phys Rev Lett 108:115701

Faber F, Lindmaa A, von Lilienfeld OA, Armiento R (2015) Crystal structure representations for machine learning models of formation energies. Int J Quant Chem 115:1094–1101

Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, Vinyals O, Kearnes S, Riley PF, von Lilienfeld OA (2017) Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. J Chem Theory Comput. 13:5255–5264. https://doi.org/10.1021/acs.jctc.7b00577

Finnis MW (2004) Interatomic forces in condensed matter. Oxford University Press, Oxford

Fujikake S, Deringer VL, Lee TH, Krynski M, Elliott SR, Csányi G (2018) Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures. J Chem Phys 148:241714

Galli G, Parrinello M (1992) Large scale electronic structure calculations. Phys Rev Lett 69:3547

Glielmo A, Sollich P, De Vita A (2017) Accurate interatomic force fields via machine learning with covariant kernels. Phys Rev B 95:214302

Glielmo A, Zeni C, De Vita A (2018) Efficient nonparametric n-body force fields from machine learning. Phys Rev B 97:184307 https://doi.org/10.1103/physrevb.97.184307

Goedecker S (1999) Linear scaling electronic structure methods. Rev Mod Phys 71:1085

Gonzalez TF (1985) Clustering to minimize the maximum intercluster distance. Theor Comput Sci 38:293

Grisafi A, Wilkins DM, Csányi G, Ceriotti M (2018) Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. Phys Rev Lett 120:036002

Haar A (1933) Der Massbegriff in der Theorie der kontinuerlichen Gruppen. Ann Math 34:147

Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-Means Clustering Algorithm. J R Stat Soc Ser C (Appl Stat) 28:100

Imbalzano G, Anelli A, Giofré D, Klees S, Behler J, Ceriotti M (2018) Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. J Chem Phys 148:241730

John ST, Csányi G (2017) Many-Body Coarse-Grained Interactions Using Gaussian Approxima-
    tion Potentials. J Phys Chem B 121:10934

Kajita S, Ohba N, Jinnouchi R, Asahi R (2017) A Universal 3D Voxel Descriptor for Solid-State
    Material Informatics with Deep Convolutional Neural Networks. Sci Rep 7:1

Kazhdan M, Funkhouser T, Rusinkiewicz S (2003) Rotation Invariant Spherical Harmonic
    Representation of 3D Shape Descriptors. In: Proceedings of the 2003 Eurographic-
    s/ACM SIGGRAPH symposium on geometry processing, SGP'03. Eurographics Association,
    pp 156–164

Mahoney MW, Drineas P (2009) CUR matrix decompositions for improved data analysis. Proc
    Natl Acad Sci USA 106:697

Morawietz T, Singraber A, Dellago C, Behler J (2016) How van der Waals interactions determine
    the unique properties of water. Proc Natl Acad Sci USA 113:8368

Nguyen TT, Szekely E, Imbalzano G, Behler J, Csányi G, Ceriotti M, Götz AW, Paesani F
    (2018) Comparison of permutationally invariant polynomials, neural networks, and Gaussian
    approximation potentials in representing water interactions through many-body expansions. J
    Chem Phys 148:241725

Prabhakaran S, Raman S, Vogt JE, Roth V (2012) Automatic Model Selection in Archetype Anal-
    ysis. In: Joint DAGM (German Association for pattern recognition) and OAGM symposium.
    Lecture Notes in Computer Science, vol 7476. Springer, Berlin Heidelberg, pp 458–467

Prodan E, Kohn W (2005) Nearsightedness of electronic matter. Proc Natl Acad Sci USA
    102:11635

Quinonero-Candela JQ, Rasmussen CE (2005) A Unifying View of Sparse Approximate Gaussian
    Process Regression. J Mach Learn Res 6:19391959

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press,
    Cambridge

Rupp M, Tkatchenko A, Müller K-R, von Lilienfeld OA (2012) Fast and accurate modeling of
    molecular atomization energies with machine learning. Phys Rev Lett 108:058301

Rowe P, Csányi G, Alfè D, Michaelides A (2018) Development of a machine learning potential for
    graphene. Phys Rev B 97:054303

Schölkopf B, Smola AJ (2002) Learning with kernels: support vector machines, regularization,
    optimization, and beyond. MIT Press, Cambridge

Schütt KT, Glawe H, Brockherde F, Sanna A, Müller KR, Gross EKU (2014) How to represent
    crystal structures for machine learning: Towards fast prediction of electronic properties. Phys
    Rev B 89:205118

Smith JS, Isayev O, Roitberg AE (2017) ANI-1: an extensible neural network potential with DFT
    accuracy at force field computational cost. Chem Sci 8:3192

Snelson E, Ghahramani Z (2006) Sparse Gaussian Processes using Pseudo-inputs. In: Weiss V,
    Schölkopf B, Platt JC (eds) Advances in neural information processing systems 18 (NIPS 2005)
    MIT Press, pp 1257–1264

Solak E, Rasmussen CE, Leith DJ, Murray-Smith R, Leithead WE (2003) Derivative observations
    in Gaussian Process Models of Dynamic Systems. In: NIPS'02: Proceedings of the 15th
    International Conference on Neural Information Processing System 2002, pp 1057–1064

Szlachta WJ, Bartók AP, Csányi G (2014) Accuracy and transferability of Gaussian approximation
    potential models for tungsten. Phys Rev B 90:104108

Thompson AP, Swiler LP, Trott CR, Foiles SM, Tucker GJ (2015) Spectral neighbor analysis
    method for automated generation of quantum-accurate interatomic potentials. J Comput Phys
    285:316

Tikhonov AN, Goncharsky A, Stepanov VV, Yagola AG (1995) Numerical methods for the solution
    of ill-posed problems. Kluwer Academic, Dordrecht

Varshalovich DA, Moskalev AN, Khersonskii VK (1988) Quantum theory of angular momentum.
    World Scientific, Singapore

Yang W (1991) Direct calculation of electron density in density-functional theory. Phys Rev Lett
    66:1438

Zhang L, Han J, Wang H, Car R, Weinan E (2018) Deep potential molecular dynamics: A scalable
    model with the accuracy of quantum mechanics. Phys Rev Lett 120:143001

# Machine Learning and Big-Data in Computational Chemistry

# 88

Rafael Gómez-Bombarelli and Alán Aspuru-Guzik

## Contents

### Abstract

Experimental chemistry and the younger discipline of computational chemistry have always aspired to increase data volume, velocity, and variety. The recent software developments in machine learning, databases and automation and hardware advances in fast co-processors, networking, and storage have boosted

R. Gómez-Bombarelli (✉)
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: rafagb@mit.edu; rgbombarelli@gmail.com

A. Aspuru-Guzik
Department of Chemistry and Department of Computer Science, University of Toronto, Toronto, ON, Canada

Vector Institute, Toronto, ON, Canada
e-mail: aspuru@chemistry.harvard.edu; alan@aspuru.com

automation and digitization. Computational chemistry is seemingly on the verge of a big-data revolution.

In this chapter, we discuss how many of these data-driven paradigms are part of long-term trend and data have long been at the heart of many chemical problems. Historical repositories of chemical data where the modern cheminformatician can mine high value curated training data are reviewed. Modern automation tools and datasets available for high-data computational chemistry are described. Current applications of computer-driven discovery of molecular materials in optoelectronics (photovoltaics and light-emitting diodes) and electrical energy storage are discussed. Finally, the impact of machine learning approaches to computational chemistry areas of structure-property relationships and chemical space, with an emphasis on generative models, are analyzed.

## 1    Introduction

Chemistry has been a *big-data* enterprise for multiple decades. Long before the term was popularized, chemists strived to wield chemical datasets whose volume, variety, and velocity exceeded the ability of tools available to them. For one, chemists have long struggled to address the immensity of chemical space. Since the arrival of computers, chemists have been creators and early adopters of digital technologies for the storage, processing, and retrieval of chemical data. Along those same lines, the advent of the Internet set off a new era of accelerating knowledge creation and sharing in the chemical sciences.

In fields such consumer analytics, advertising, transportation, healthcare or finance, the clearest gains from big-data, and machine learning approaches have come from subareas that combine as many favorable features as possible: data is inexpensive to generate, digital in origin or easy to digitize, highly organized and consistent, and extremely abundant. While abundant, chemical data tends to be expensive to generate, heterogeneous, hand-made, and sparse. Chemical data entry and curation is generally done manually; these tasks require significant expertise because of the heterogeneous sources and reporting criteria, and the technical complexity of chemical reports. In recent years, as chemical instruments, such as microscopes or spectrometers, produce digitized data, very large datasets can be generated by a single device. This very much opens the door to leverage *local* big data, but one still faces similar barriers when attempting to consolidate results from different groups or apparatuses.

Computational chemistry approaches share a series of features that make them very amenable to this data-driven paradigm and are proving to be the perfect testing ground for realizing the big-data vision in chemistry. The output of computational chemistry software is digitized from beginning to end. Even if it relies on custom-formatted input and output files, automating the creation and processing of these files is a simple task. Computational results are replicable to floating point precision when using the same computer code, and with extreme accuracy even among different programs as long as they use the same underlying algorithms. Because compute

is essentially a commodity, results can be generated in arbitrary amounts with linear scaling in cost – or even sublinear, thanks to machine learning approaches. Finally, there is a well-understood tradeoff between computational cost and accuracy that can be leveraged through data-driven techniques.

In this chapter, we will analyze how computational chemistry has embraced big-data for accelerated discovery of much needed new molecules and how these newly generated datasets have allowed to apply, and develop, machine learning techniques that make them even more powerful.

## 2 Repositories of Chemical Data

The initial development of big-data approaches in chemistry was historically focused on consolidating, curating, and indexing literature and empirical data. Tackling these tasks was one of the first applications of computers in chemistry and remains a key driver in the development of new computational tools for chemistry.

The Chemical Abstracts Service (CAS) was created in 1907 to keep complete records of the chemical literature and reported substances, following the idea of previous abstracts journals in Europe such as *Chemisches Zentralblatt* and nineteenth-century compilations such as *Beilsteins Handbuch der organischen Chemie*, *Gmelins Handbuch der anorganischen Chemie*, or the Merck Index. It contained over 12,000 indices in its first year, its one millionth record in 1937. CAS historically relied on thousands of volunteers to create abstracts and index information for each paper – not that far from today's Mechanical Turks at Amazon – and slowly phased them out between the late sixties and the nineties. In addition, CAS employed hundreds of staff members by the 1950s, underscoring both the great value of big data to chemistry and the great cost of getting that same data into neatly formatted into useable form.

The adoption of digital computers for chemical applications in the late 1960s initiated the digitization of chemical big data by facilitating the storage, entry, and very importantly querying of chemical databanks. CAS computer-based Chemical Registry System, like the Beilstein Registry Number before it, assigns unique individual arbitrary numbers to chemicals. It debuted internally in 1964 and was progressively updated and expanded, making the basis for increasingly modern querying and retrieval tools, using command-line and graphical interfaces to retrieve data from local or online records. With around three million records and growing at hundreds of thousands per year in the mid-seventies, the registry grew exponentially to its 30 millionth chemical substances in 2007, 50 millionth in 2009, and 100 millionth in 2015. Competing commercial efforts that also build on older academic work are Elservier's Reaxys (Beilstein and Gmelin) and Thomson-Reuters, now Clarivate.

In 1965, the Kennard group at Cambridge University started collecting published crystallographic data for small molecules. This effort grew into what is known as the Cambridge Structural Database, which now hosts nearly one million curated entries from x-ray and neutron diffraction analyses (Groom et al. 2016). Also in the

late 1960s, joint efforts between Brookhaven National Laboratory and Texas A&M University initiated what would become the Protein Data Bank, a repository for three-dimensional structural data of large biological molecules, such as proteins and nucleic acids, and now holds over 120,000 entries (Berman et al. 2000; Bernstein et al. 1978; Meyer 1997).

Other repositories of experimental crystallographic information (Bruno et al. 2017) are the Crystallography Open Database (COD) (Gražulis et al. 2009, 2012), which stores user-submitted data; the Inorganic Crystal Structure Database, produced cooperatively by FIZ Karlsruhe and NIST that stores crystal structures of inorganic solids (Belsky et al. 2002); and the International Centre for Diffraction Data (ICDD) that stores powder diffraction patterns (Faber et al. 2002).

The rise of the Internet sparked the launch of other big-data repositories and the expansion of online availability of existing ones both in chemistry and materials. The Registry of Research Data Repositories aggregates lists of multiple repositories and datasets along many scientific disciplines including chemistry and materials science (Pampel et al. 2013).

The National Institute of Standards and Technology has hosted the NIST Chemistry WebBook since the 1996 (Linstrom and Mallard 2001), a compendium of mostly spectroscopic and thermodynamic data originally compiled in handbooks and tables. In addition, the nearly 200 Standard Reference Data sets include other data of interest to chemists. NIST 101 (III 1999) in particular contains calculated data quantum chemical data for about 1800 gas-phase atoms and small molecules and tools for comparing experimental and computational ideal-gas thermochemical properties.

Since patents and patent applications are open, they are a big-data, information-rich asset. Patent documents, however, tend to be only text-based or image-based, and their digitization into a format that is useable for data-driven approaches is a standing challenge. The European, United States, Japanese Patent Offices, and the World Intellectual Property Organization can be accessed and queried online, but they are not systematically annotated. The US Patent and Trademark Office, through their Complex Work Unit program, made available digital representation of molecules in patent applications, easing the data-processing pipelines.

In the open-data arena, NextMoveSoftware released close to million chemical reactions, extracted by means of automated text mining of the relevant experimental sections reported in patents, covering the period between 1976 and 2013. SCRIPDB contains curated syntheses, chemicals, and reactions from the patent literature, collected from CWU files coming from granted US patents (Heifets and Jurisica 2012).

SureChem is another molecule database, open-sourced in 2012 to SureChemBL. It used by IBM's Strategic IP Insight Platform and initially released more than two million chemical structures extracted from about 4.7 million patents (1976–2000, only text) and subsequently extended to include patents published up to the end of 2010 and chemical structures were additionally derived from US CWUs and images (2001–2010) (Papadatos et al. 2016). UniChem attempts to unify and cross reference these multiple databases (Chambers et al. 2013).

ChemSpider is a database of chemicals with more than 60 million entries from over 480 data sources. Originally a private enterprise, it was acquired by the Royal Society of Chemistry (RSC) in 2009. It includes a crowd-sourced component, but only limited downloads are available (Pence and Williams 2010).

Because of the large size of the set of possible small organic molecules with biological activity, the large number and diversity of biological targets, the high dimension of many biological assays and measurements and the high value of healthcare applications, cheminformatics, biological chemistry, and drug discovery applications are key target for big data applications in chemistry. Toxicological information of drugs and chemicals are also interesting biological interactions that are recorded in several publicly available databases.

The National Center for Biotechnology Information (NCBI), part of the United States National Library of Medicine (NLM) from the National Institutes of Health (NIH), hosts multiple datasets and informatics tools in biotechnology. Of particular interest in chemistry is PubChem, a database of molecules and their activities against biological assays. Initiated in 2004, it now aggregates over 540 sources, over 90 million compounds, and 233 million bioactivity results for nearly 2.5 million of those. As a federally sponsored service, PubChem has been seen to be in conflict with for-profit repositories such as CAS (Kaiser 2005).

ChemBL is an open database that contains binding, functional, and ADMET (absorption, distribution, metabolism, excretion) measurements for drug-like biologically active compounds, with nearly 15 million bioactivity measurements for more than 1 million compounds and 11,500 protein targets (Gaulton et al. 2012). DrugBank is a web-accessible cheminformatics database and service combining structural and biological target data for drug molecules. The database contains in excess of 9000 small molecule drugs, 3000 FDA-approved drugs, and data for nearly 17,000 drug-target associations (Wishart et al. 2006, 2018). The database for Chemical Entities of Biological Interest follows a similar focus as well (Degtyarenko et al. 2008).

The PDBbind database matches published affinity constants from the literature for the ligand-protein systems whose 3D structures are stored in PDB, undergoing periodic updates (Liu et al. 2015; Wang et al. 2004). BindingDB (Chen et al. 2001; Gilson et al. 2016) and AffinDB (Block et al. 2006) also aim to fulfill a similar task. The comparative Toxicogenomics Database is a volunteer-based genomic resource devoted to toxicologically relevant genes and proteins and their interactions with chemicals and toxins (Davis et al. 2017; Mattingly et al. 2003). MACiE contains enzyme reaction mechanisms focused on the evolution of enzyme catalytic mechanisms and the classification with respect to chemical mechanism (Holliday et al. 2005).

These services are big-data pioneers in chemistry and highlight both the early understanding of how chemistry is a high-dimensional, sparse-data arena where big-data approaches can create great value, as well as the high cost of gathering and curating chemical data. For machine learning applications, these repositories provide extremely valuable labeled data.

# 3    Big Computational Chemistry

The field of computational chemistry quickly followed the deployment of the first digital computers and has grown at a fast pace, matching developments in algorithms and hardware and also contributing its own. Oftentimes, the chemical calculations carried out will push the hardware available to run them: running in parallel over many cores in supercomputers, using large amounts of memory or storage.

In that sense, judging by the volume and velocity of the output data they produce, and by the strain on hardware requirements, computational chemistry calculations have been big-data all along, essentially as big as the available resources allowed and have been drivers of big-data technologies.

In this chapter, we will focus on applications of computational chemistry that combine volume and velocity with also high variety because of a large degree of granularity. These are the cases of (i) distributive computing, where small calculation payloads are distributed to a large grid of small computers, and after computation, the results are consolidated and processed, and (ii) high-throughput virtual screening, where many thousands of candidate molecules or materials are calculated individually in an automated fashion and a data-driven search for the most performant candidate materials is carried out.

## 3.1    Automation and Databases in Computational Chemistry

Many tools have been developed to automate the creation, submission, transferring, processing, parsing, storage, and querying of computational chemistry data. One of the most venerable web-based examples is the Basis Set Exchange, originally assembled at the Environmental Molecular Sciences Laboratory, where a myriad of curated basis sets for most of the periodic table are available for download in multiple formats (Feller 1996; Schuchardt et al. 2007).

Close to half dozen platforms exist with a similar philosophy towards achieving some or all the following: automation of materials science and solid-state electronic structure calculations; data processing and analysis of those calculations; and centralized, web-accessible repositories of the output of these calculations for virtual discovery and machine learning purposes. The multiple solutions offer somewhat overlapping functionality, generally in the materials space, and have been reviewed recently (Lin 2015).

The Electronic Structure Project (Klintenberg et al. 2002; Ortiz et al. 2009) utilized the structural data from the ICSD to screen for novel inorganic materials.

The Computational Materials Repository (Landis et al. 2012) proposes an integrated software solution for computer-driven materials design. It is part of the ecosystem of the Quantum Materials Informatics Project that also includes the Atomic Simulation Environment, a python library for working on atomistic simulations (Hjorth Larsen et al. 2017).

AFLOW is an automatic framework for high-throughput materials discovery (Calderon et al. 2015; Curtarolo et al. 2012a), and the matching repository Aflowlib.org hosts and serves the results of those calculations to the public (Curtarolo et al. 2012b). It now contains 1,748,704 material compounds – with over 173,121,696 calculated properties.

The Python Materials Genomics (pymatgen) is an open-source python library for materials for the analysis of solid-state DFT calculations (Ong et al. 2013) which also has a matching Materials Application Programming Interface (Ong et al. 2015) to interact with the Materials Project, a large-scale database of materials calculations (Jain et al. 2013). Other computational tools in this ecosystem include tools such as Custodian for error handling, Fireworks for workflow management. The Materials Project tools and data have been used in over 100 published papers, and nearly 200 by its creators.

The Open Quantum Materials Database (OQMD) is a fully open project that hosts over 400,000 DFT energy calculations of compounds from the ICSD and also for hypothetical compounds, potentially uncovering valid, but yet to synthesize chemistries (Kirklin et al. 2015).

AiiDA (Pizzi et al. 2016) is a flexible and scalable informatics' infrastructure for simulations, data, and workflows with a heavy focus on plane-wave DFT calculations of materials and much attention to data provenance (Merkys et al. 2017).

The Novel Materials Discovery repository was established to host, organize, and share materials data in a pipeline-agnostic way (Goldsmith 2016) and hosts over 44 million open access user-submitted total-energy calculations from a variety of computer codes.

The ioChem-BD Platform provides a similar solution with an emphasis on molecular data (Álvarez-Moreno et al. 2015). Several toolkits address the simulation and role of defects, such as MAST (Mayeshiba et al. 2017), PyDII (Ding et al. 2015), and others (Goyal et al. 2017). PyChemia is a python library for automatize atomistic simulations, with a focus on materials and interfaces to some DFT codes and data mining functionality.

Multiples datasets of computational chemistry results aimed purely at generating diverse training data exist. QM9 contains B3LYP/6–31 G(2df,p) results for 134k stable small organic molecules made up of CHONF, including harmonic frequencies (Ramakrishnan et al. 2014). PubChemQC is a recent attempt to create training data for machine learning approaches that calculated the ground-state electronic structures of three million molecules based on density functional theory (DFT) at the B3LYP/6-31G* level and 10 lowest excited states of over two million molecules at TD-DFT/B3LYP/6–31 + G* level of theory (Nakata and Shimazaki 2017). ANI-1 contains energies and DFT-level properties for 20 million conformations for over 50,000 small organic molecules distorted along normal modes (Smith et al. 2017). Ab initio molecular dynamics are also available: 5,000 frames at 500 K at the PBE level of theory for 113 structural isomers of $C_7O_2H_{10}$ and hundreds of thousands frames for 8 small organic molecules (Chmiela et al. 2017; Schütt et al. 2017).

A large number of internal tools for data processing have grown into open libraries that are available for data analysis of calculation outputs: the python-based

cclib for parsing and interpreting the results of computational chemistry packages (O'boyle et al. 2008); ESTEST, a free framework for the comparison, validation and sharing of quantum chemical calculation outputs (Yuan and Gygi 2010); ORBKIT, also python-based library for postprocessing of quantum chemical wavefunction outputs from multiple codes; (Hermann et al. 2016); and PyGlobal, spreadsheet-oriented output postprocessing tool for DFT calculations (Nath et al. 2016). Other tools are also available for input generation and pipelining calculations, such as JACOB (Waller et al. 2013) a framework for computational chemistry aimed at enterprise application, PyADF for scripting multiscale quantum chemistry using the ADF package (Jacob et al. 2011).

This ample landscape suggests that increasing both the velocity and volume of quantum chemical calculations is of great interest. These increases, however, come with a tradeoff in the variety of the applications. Because of the added complexity, most of these rarely see adoption outside the groups or consortia that created them (Thygesen and Jacobsen 2016).

## 3.2 High-Throughput Virtual Screening

As computational methods become more accurate and computing hardware more affordable, the possibility of automatically prescreening compounds virtually before synthesis grows more promising. Various teams have used HTVS for discovery of many inorganic materials. These are treated in depth in other chapters of this book. Here, the focus will be on organic molecular materials in the domains of organic optoelectronics for light-energy interconversion (photovoltaics and light emitting diodes) and for electrical energy storage (Pyzer-Knapp et al. 2015).

### 3.2.1 Optoelectronics

Merging concepts from both the volunteer, distributive computing efforts and the HTVS vision, lies the Clear Energy Project (CEP). Since its inception around 2006 and throughout two phases, this project ran on IBM World Community Grid, where volunteers donated computer time to virtually screen *p*-type, and later *n*-type, organic photovoltaic oligomers.

CEP tackled many of the challenges for automated virtual testing of chemical compounds: programmatic generation of candidate molecules, automation of the quantum chemical calculation, data storage, and analysis.

Combining a pool of 20 fragments through covalent linking at active sites and also by forming fused rings adjacent to these labeled reactive atoms, the CEP molecular generation processed went up to tetramers and produced over two million candidates that were screened exhaustively. The quantum chemical calculations carried out included multiple DFT functionals and amount to the largest computational chemistry project to that date. The candidate tetramers were assessed using the Scharber model for photovoltaic efficiency, assuming fullerenes as electron acceptors and both raw and empirically calibrated donor energy levels. The statistical analysis of the large dataset produced afforded correlations for the

most and least promising fragments and fragment combinations (Hachmann et al. 2011, 2014; Olivares-Amaya et al. 2011).

The CEP, because of the large theoretical dataset, and the existence of numerous independent experimental results in the area of OPV has proven a testing ground for novel approaches to materials discovery. A probabilistic kernel-based calibration scheme to improve theoretical gas-phase results and to capture the bulk effects using a collection of experimental results from the literature (Lopez et al. 2016) improved the predictive performance of DFT calculations (Pyzer-Knapp et al. 2016).

A later subproject has focused on the virtual screening of over 50,000 non-fullerene electron acceptors from the combination of over 100 common organic moieties. Time-dependent density functional theory calculations were also carried out for elected lead compounds. Diketopyrrolopyrroles and quinoidal thiophene derivatives showed good promise and were proposed for additional study (Lopez et al. 2017).

At a less gargantuan scale, other works also addressed screening over bulk heterojunction solar cell components, such as combinatorial band-gap design strategy over 780 different copolymer donor materials (Shin et al. 2014).

Some works have addressed virtual discovery of molecules for other classes of solar cells, such as $TiO_2$-based dye-sensitized solar cells, optimizing over common dyes (Martsinovich and Troisi 2011) and also over porphyrins (Ørnsø et al. 2014).

A successful example of blending computational chemistry with deep learning for applied materials discovery has been reported in the area of thermally activated delayed fluorescence (TADF) organic light-emitting diodes (Gómez-Bombarelli et al. 2016). Using custom software that mimics cross-coupling reactions on existing staring materials, a database spanning nearly two million feasible compounds was created from over two hundred starting donor and acceptor fragments. The TADF character of the compounds, as well as their color, was estimated using accurate, empirically calibrated TD-DFT calculations. For accelerated results, the candidates were screened through a neural network, using topological fingerprints as features. The leading compounds with the most promising predicted chemical properties were assessed by a team of experts who synthesized and tested the consensus champion compounds. These were then tested in optoelectronic devices, where they matched the performance of human-generated champion compounds (Fig. 1).

### 3.2.2   Electrolytes and Energy Storage

Energy storage is one of the most active areas of materials science and engineering, given the strong demand in both lightweight, high-energy density applications such as mobile phones or transportation and static, low-cost, grid-scale storage.

Flow batteries are large, static batteries where liquid electrolytes are stored in tanks and circulated across an electrochemical cell when charging or discharging. Although they show somewhat inferior energy densities compared to solid-state batteries such as lithium-ion, flow batteries are potentially a much better solution to grid-scale electrical storage because of the lower cost and use of earth-abundant and cheap electrolyte materials. The independent scaling of power (depending on effects such as the kinetics of the electrochemical reaction and the electrode surface area)
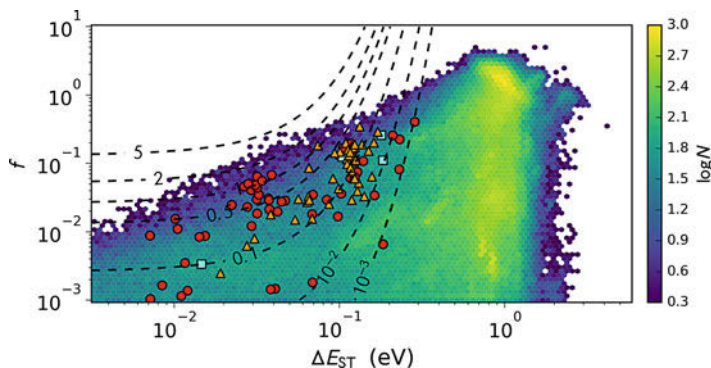
**Fig. 1** Overview of TADF OLED chemical space. Red dots represent overlap with previous experimental reports in the literature, orange triangles correspond to theoretical leads, cyan squares represent experimentally confirmed theoretical leads. The log-density of molecular candidates as a function of singlet-triplet gap and oscillator strength of shown. The dashed lines follow iso-contours of estimated rate of delayed fluorescence

and energy (related to the total size of the tanks, the concentration of the electrolyte in solution, and the degree of ionization of the electrolyte) allows more flexible engineering for flow-battery solutions. Using water as solvent further reduces the installation costs and increases the operational safety. Because they can be sources from oil and from sustainable biomatter, the great design flexibility of organic molecules, and their promising performance, organic electrolytes for flow batteries are heavily considered contestants (Kowalski et al. 2016; Leung et al. 2017; Wei et al. 2017).

High-throughput simulation, particularly when coupled with experimental follow-on, has become a powerful tool in the discovery of electrolyte materials for organic flow batteries. The report of an efficient, metal-free, low-cost flow battery design using acidic aqueous solutions of anthraquinone disulfonate and bromine/hydrobromic acid leveraged thousands of DFT calculations of potential quinone molecules to optimize reduction potential of the negolyte and its aqueous solubility (Huskinson et al. 2014). The careful analysis of the theoretical predictions allowed to identify useful design rules and structure-property relationships over the domain of benzo-, naphto-, and anthraquinones bearing any of 13 functional groups (Er et al. 2015). Other high-volume calculations of bio-inspired thiophenoquinone derivatives also identified potentially useful electrolytes for flow batteries (Pineda Flores et al. 2015).

The Electrolyte Genome project used the Materials Project backbone to perform high-throughput calculations on a set of nearly 5000 molecules derived mostly from quinoxaline and anthraquinone, thiane, thiophene, and bipyridine derivatives (Qu et al. 2015). Special attention was paid to error handling and estimation of redox potentials, ion pair dissociation, and complex salt formation. A detailed analysis of nearly 1400 quinoxalines for nonaqueous flow batteries estimating redox potentials,

solvation energies, and structural changes was also reported (Cheng et al. 2015), as well as over 4000 five- or six-membered rings with one or two functional groups attached (Pelzer et al. 2017).

Somewhat smaller scale virtual searches have also been reported. Starting from experimentally reported molecules available in PubChem, and down-selecting manually to 315 after deploying a neural network, improved experimental performance was observed for a quinoxaline derivative (Park et al. 2015, 2016).

The chemical stability of organic molecules for electrochemical applications, which impact battery shelf-life and long-term cyclability, is one of the standing issues of virtual discovery applications. For this purpose, the first and second reduction and oxidation of organic solvents for lithium-ion batteries have been analyzed using DFT (Borodin et al. 2015). Similarly, bio-inspired alloxazine electrolytes have been shown experimentally to provide excellent stability in alkaline aqueous flow batteries, by including stability criteria in the virtual search, in addition to redox and solubility requirements (Lin et al. 2016).

Smaller approaches with less focus on automation and extreme throughput have been reported, focusing on aspects such as low $Li^+$ binding affinity over 32 organic molecules (Park et al. 2011), 15 counterions for magnesium electrolytes (Qu et al. 2017), redox-switchable polymer-based membranes (Ward et al. 2017), or tuning the reduction potential of organic molecules to optimize the combination of small redox active molecules with conducting polymers (Araujo et al. 2017).

HTVS has also been applied for materials related to other electrochemical applications such as double layer capacitors (Schütter et al. 2016) or other green-chemistry applications such as switchable-hydrophilicity solvents (R. Vanderveen et al. 2015).

## 4 Applied Machine Learning for Accelerated Discovery

As is the case in many other areas, machine learning applications to chemistry are not a novel enterprise (Pierce and Hohne 1986). As is also the case in many other applications, the combination of deep learning algorithms with larger datasets and specialized computing hardware has resulted in many effective applications of machine learning for chemistry.

## 4.1 Quantitative Structure-Property Relationships

Most applications of machine learning to computational chemistry applications are related to building fast statistical proxies to expensive calculations on three-dimensional atomic arrangements or chemical graphs. This is, in a sense, building quantitative structure-property relationships trained either on molecular structures or on 3D geometries, often generated on-the-fly using molecular mechanics.

Other chapters in this same publication give a thorough review of many recent contributions to this area, so only a brief overview will be provided. Readers are encouraged to explore further in this same book.

The idea of using neural network models to interpolate over potential energy surfaces goes back to the 1990s (Blank et al. 1995). Very early contemporary works applying machine learning to computational chemistry are those by Behler where neural networks were used for molecular dynamics simulations and other PES exploration techniques (Behler 2011b, a; Behler et al. 2007; Behler and Parrinello 2007). These and many other applications of neural-networks, mostly to bulk and heterogenous systems, have been reviewed recently (Behler 2017).

Another line of application of machine learning techniques to computational chemistry, with a heavier focus on molecules, rather than bulk materials initially leveraged kernel methods, with emphasis on mapping 3D coordinates (Rupp et al. 2012; Schütt et al. 2017) or electronic densities (Brockherde et al. 2017; Snyder et al. 2012) to properties. These approaches are well covered in other chapters of this publication.

One of the most important issues in predictions based on molecular structures is the choice of features to represent molecules. Chemists tend to think of molecules as undirected graphs, with atoms in the nodes and bonds in the edges, and some additional considerations for stereochemistry. In quantum chemistry, molecules are three-dimensional arrangements of atoms that correspond to a local minimum in the potential energy hypersurface. For multiple decades, cheminformatics applications, mostly in pharma and drug discovery, have created descriptive, rich features for machine learning over both graphs: pharmacophore fingerprints, topological fingerprints, and also 3D conformations.

Recent applications of machine learning to molecules in chemistry have focused on QSPR over quantum-chemical properties and placed their focus on more information-rich features. The Coulomb matrix, for instance, is a permutation-dependent symmetry invariant modified distance matrix. Contemporary highly powerful neural network approaches also leverage 3D information (Gilmer et al. 2017; Lubbers et al. 2017; Schütt et al. 2017). In the case of molecules whose 3D structure is unknown, which is generally the case when carrying out predictions, these features are calculated from guess geometries calculated on the fly at the molecular mechanics level of theory. Molecular mechanics methods, through the underlying force fields, are heavily parametrized, generally on quantum chemical calculations but eventually on the experimental data used to parametrize those too. Therefore, the machine learning methods applied over these features are heavily supervised, with their inputs essentially embedding quantitatively a large amount of prior chemical knowledge and hand-tuning.

Learning on graphs presents a very unique set of challenges, because of graph-isomorphism. Machine learning approaches over graph structures have much interest beyond chemistry, such as in networks of any kind in transportation, databases, telecommunications. Whereas extended connectivity circular fingerprints address graph isomorphism, the process of encoding the graph, hashing, muddles chemical information and can lead to different chemical substructures activating the

same feature. Neural fingerprints, a differentiable deep-learning extension of ECFP, have recently been proposed as a more flexible and learnable alternative (Duvenaud et al. 2015). These continuous counterparts of topological fingerprints and derivative graph convolutions for chemistry (Kearnes et al. 2016) have proven to outperform some traditional cheminformatics descriptors from QSPR, particularly for larger datasets, and open a new avenue for molecular screening.

## 4.2    Searching and Optimizing in Chemical Space

An interesting subset of applications of machine learning to the computational chemistry/cheminformatics community is moving in the discrete, high-dimensional space spanned by all possible molecules, or by a relevant subset, such as the small organic molecules with aromatic rings, or similar subsets of interest.

Chemical space is extremely large, with estimates ranging by many orders or magnitude, from $10^{23}$ to $10^{60}$. Rule-based efforts to enumerate molecules, exhaustively for small compounds and culling nonpractical but formally valid and even potentially stable molecules, have been reported as follows: GDB-11 lists 26.4 million small organic molecules of up to 11 atoms of C, N, O, and F (Fink and Reymond 2007); GDB-13 enumerates 978 billion molecules with up to up to 13 atoms of C, N, O, S, and Cl (Blum and Reymond 2009); and GDB-17 contains 166.4 billion molecules of up to 17 atoms of C, N, O, S, and halogens (Ruddigkeit et al. 2012). Extrapolating from those sets, the size of drug-like chemical space has been extrapolated to $10^{33}$ (Polishchuk et al. 2013).

In addition to the size of the space, molecules are discrete graphs and rules exist regarding the types and degrees of connectivity that are allowed. Even further, molecules that are formally valid may still be chemically unstable at the temperatures, pressures, and timescales of interest. Hence, performing local, or even more challenging global, optimization is of big interest in materials and drug design and also a great challenge. The Chemical Space Project leverages the GDB databases to develop visualization and exploration tools, with a focus on drug candidate molecules (Reymond 2015).

Custom molecular libraries have proven a very effective way of navigating chemical space, as the human-driven design allows to embed rules and chemical requirements that are hard to capture without strong supervision, particularly regarding synthetic accessibility of molecules. This success is reflected in the ease of experimental applicability of the examples in the previous section.

Multiple machine-learning approaches to this area suggest that the large existing datasets of chemical reactions from the patent literature (Lowe 2012) or from commercial databases can be leveraged to automatically construct reactivity prediction tools for organic synthesis reactions and retrosynthetic analysis software. Recent examples include prediction the outcome of chemical reaction from fingerprints of the reactants after training on rule-generated examples (Wei et al. 2016) and selection of the major product by ranking a self-generated list of candidates (Coley et al. 2017a; Jin et al. 2017). Sequence-to-sequence models following approaches

to machine translation (Fooshee et al. 2018; Schwaller et al. 2018) or similarity searches (Coley et al. 2017b) have proven effective at automatically predicting organic synthesis reactions in an automatic, data-driven way. Even more promising results have been obtained from larger, better-curated datasets (Segler et al. 2017; Segler and Waller 2017a, b).

A well-understood approach to chemical optimization are genetic-algorithms, where mutations from a hand-prepared list of chemical transformations and stochastically applied to starting molecules, and the resulting compounds, if the show improvement in a desired property, are kept for further evolution. Because they rely on hand-picked mutations and hyperparameters, and oftentimes also on manually tuned tradeoffs between target properties, these approaches involve a certain degree of chemical intuition. By compounding mutations that may be allowed individually but not in combination, the molecules generated tend to be quite complex. These types of genetic approaches have been used to navigate chemical space (van Deursen and Reymond 2007; Virshup et al. 2013) and to optimize multiple classes of molecules, such as organic light-emitting diodes (Rupakheti et al. 2016; Shu and Levine 2015), organic photovoltaics (Kanal and Hutchison 2017), diamondoids (Teunissen et al. 2017), visible chromophores with high hyperpolarizability (Elward and Rinderspacher 2015), or small molecules with high electrical dipole (Rupakheti et al. 2015).

## 4.3    Generative Models

Generative models are machine learning models that aim to produce natural-seeming data that capture the intrinsic statistical properties of the training populations. They can be trained in an unsupervised way and thus are not inhered by the need for labeled data. Given the large size of chemical space, and the very abundant number of chemicals known (see Sect. 2) numbering around 100 million known compounds, unsupervised deep learning models based on existing molecules have been assessed recently.

One of the most basic examples of generative models in chemistry has been the use of recurrent neural networks (RNN) to predict the next character of the SMILES representation of molecules. By feeding $n$ characters to the network to predict the $n + 1$th, RNNs have proven very powerful at generating valid – if nonsensical – text (Karpathy 2015). A simple and comprehensive string representation of molecules that is human readable and can be stored a single string exists: the simplified molecular-input line-entry system (SMILES) (Weininger 1988). SMILES contain the full molecular connectivity using a series of rules and a canonicalization procedure. The same principles shown for text have proven to work well for generating molecules through their string-based SMILES representation. Additional work has been directed towards generating molecules for generating leads that capture the statistical behavior of the training data (Bjerrum 2017; Ertl et al. 2017; Gupta et al. 2018). A further step involves evolutionary refinement of the molecule pool to bias the set towards a given property (Olivecrona et al. 2017).

Autoencoders are deep learning models that are trained to reproduce a high-dimensional input, subjecting it to a low-dimension information bottleneck. Some of the thorniest issues in molecular design emerge from the discrete nature of molecules and the extremely large size of the search domain. Hence, this low-dimension, real-valued, decodable embedding would potentially allow to apply gradient-based optimization algorithms to molecular optimization. A further improvement is the use of a variational autoencoder, where random noise is added to the encoding step, which results in more continuous coverage in the latent space and fewer *dead areas* that do not correspond to valid decoded points.

As discussed above, there are available strategies for using chemical graphs as inputs for machine learning systems. However, deep learning tools that efficiently write chemical graphs are still an open problem. Neural networks architectures have been shown to efficiently write text and images, and hence, the first report (Gómez-Bombarelli et al. 2018) of variational autoencoders (VAE) for chemical discovery used a string-based chemical representation. This work showed how a VAE can accurately reconstruct molecules from a continuous real-valued array representation. Even further, it explored how transformations in the latent space, particularly molecular optimization with respect to properties of interest, can be carried out. This application was particularly efficient in the case of jointly trained VAE plus predictor systems, where the deep learning system was simultaneously trained as an unsupervised generator and a supervised predictor. In this case, the latent space is topologically shaped by chemical property (Fig. 2).

The original VAE model suffers from several flaws, generally related to its ability to write out molecules, particularly the string output is probabilistic. Because of the string representation used, and its need for internal consistency and even arithmetic (opening and closing rings and branches), the VAE models produce many invalid molecules, in more than one sense. On the one hand, syntax errors result in strings that are invalid as SMILES and do not correspond to an actual chemical graph (cycles that open but do not close, parenthesis that open but do not close, representing unfinished branches, etc.), a similar type of failure is to write complete graphs that are not chemically valid graph (generally related to valences and the octet rule: carbon atoms with valence higher than five, oxygen atoms with valence three, etc.). A different type of writeout failure is molecules that are formally and chemically valid, but when re-encoded, do not correspond to the original point in the latent space: the stochastic text generation strays away from the original point.

A number of works are rapidly expanding this area, addressing these performance issues, and exploring further avenues for deep generative models in chemistry (Blaschke et al. 2018; Xu et al. 2017). These include using a series of grammar rules for SMILES as the output of the decoder (Kusner et al. 2017), active learning over the validity of the output (Janz et al. 2017), performing constrained Bayesian optimization to avoid exploring dead areas of the latent space (Griffiths and Hernández-Lobato 2017), performing local optimization near encoded latent points (Mueller et al. 2017), combining an additional RNN to generate higher-quality outputs through reinforcement learning (Jaques et al. 2016).
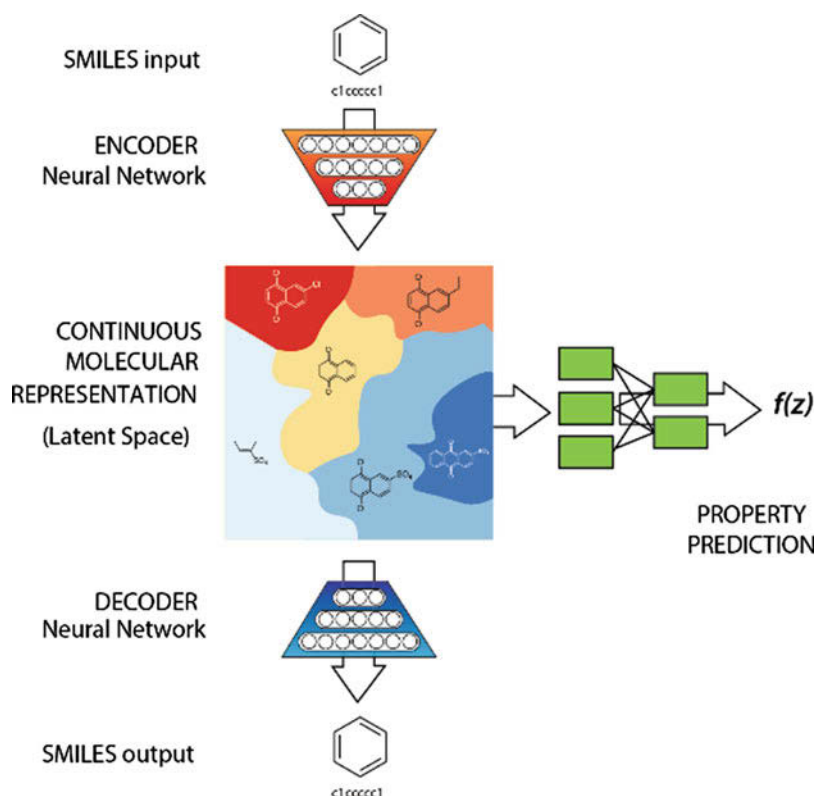
**Fig. 2** Diagram of a Variational Autoencoder with joint property prediction: an encoder neural network takes in the string repetition of a molecule and converts it into a real-valued array that can be decoded back into a string through decoded neural network. By training jointly with a property prediction network, the latent space can be navigated for property optimization

Further improvements and applications towards generating DNA (Killoran et al. 2017) and protein sequences (Sinai et al. 2017) have been reported, as well as adversarial approaches with the ability to apply bias towards desired properties (Guimaraes et al. 2017; Sanchez-Lengeling et al. 2017).

## 5    Conclusions

Machine learning and data-driven application have taken computational chemistry by storm. Tasks that have been deemed *holy grails* for decades seem closer than ever thanks to deep learning approaches: affordable computational predictions that match or surpass experimental accuracy, computerized retrosynthesis that can beat humans, etc.

Chemistry in general and computational chemistry, in particular, have long strived with high-volume, high-velocity, and high-variety data. The pain points are dissimilar however. In computational chemistry, the acquisition velocity is less and less of a bottleneck. Experimentally they are many more constraints on the pace and reproducibility, and they come at a much larger cost. Computational chemistry, on the other hand, has shown difficulty addressing variety: as parametrization increases the transferability and the trust in the predictive power decrease.

The more digitized and automated it is, the more experimental chemistry assimilates to computational chemistry, and hence the easiest to leverage these extremely promising data-driven tools. For a broader impact, and if computational chemistry is truly a sandbox and an accelerator for ideas that will ultimately change experimental chemistry, more focus is needed on addressing heterogeneous, unstructured data. Unsupervised machine learning and transfer learning are promising tools for this task.

# References

Álvarez-Moreno M, de Graaf C, López N, Maseras F, Poblet JM, Bo C (2015) Managing the computational chemistry big data problem: the ioChem-BD platform. J Chem Inf Model 55:95

Araujo RB, Banerjee A, Panigrahi P, Yang L, Strømme M, Sjödin M, Araujo CM, Ahuja R (2017) Designing strategies to tune reduction potential of organic molecules for sustainable high capacity battery application. J Mater Chem A 5:4430

Behler J (2011a) Atom-centered symmetry functions for constructing high-dimensional neural network potentials. J Chem Phys 134:74106

Behler J (2011b) Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. Phys Chem Chem Phys 13:17930

Behler J (2017) First principles neural network potentials for reactive simulations of large molecular and condensed systems. Angew Chemie Int Ed 56:12828

Behler J, Lorenz S, Reuter K (2007) Representing molecule-surface interactions with symmetry-adapted neural networks. J Chem Phys 127:14705

Behler J, Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett 98:146401

Belsky A, Hellenbrandt M, Karen VL, Luksch P (2002) New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. Acta Crystallogr Sect B Struct Sci 58:364

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1978) The protein data bank: a computer-based archival file for macromolecular structures. Arch Biochem Biophys 185:584

Bjerrum EJ (2017) SMILES enumeration as data augmentation for neural Network modeling of molecules arXiv:1703.07076. https://arxiv.org/abs/1703.07076

Blank TB, Brown SD, Calhoun AW, Doren DJ (1995) Neural network models of potential energy surfaces. J Chem Phys 103:4129

Blaschke T, Olivecrona M, Engkvist O, Bajorath J, Chen H (2018) Application of generative autoencoder in de Novo molecular design Mol. Inform 37:1700123

Block P, Sotriffer CA, Dramburg I, Klebe G (2006) AffinDB: a freely accessible database of affinities for protein-ligand complexes from the PDB. Nucleic Acids Res 34:D522

Blum LC, Reymond J-L (2009) 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. J Am Chem Soc 131:8732

Borodin O, Olguin M, Spear CE, Leiter KW, Knap J (2015) Towards high throughput screening of electrochemical stability of battery electrolytes. Nanotechnology 26:354003

Brockherde F, Vogt L, Li L, Tuckerman ME, Burke K, Müller K-R (2017) Bypassing the Kohn-Sham equations with machine learning. Nat Commun 8:872

Bruno I, Gražulis S, Helliwell JR, Kabekkodu SN, McMahon B, Westbrook J (2017) Crystallography and databases. Data Sci J 16:38

Calderon CE, Plata JJ, Toher C, Oses C, Levy O, Fornari M, Natan A, Mehl MJ, Hart G, Buongiorno Nardelli M, Curtarolo S (2015) The AFLOW standard for high-throughput materials science calculations. Comput Mater Sci 108:233

Chambers J, Davies M, Gaulton A, Hersey A, Velankar S, Petryszak R, Hastings J, Bellis L, McGlinchey S, Overington JP (2013) UniChem: a unified chemical structure cross-referencing and identifier tracking system. J Cheminform 5:3

Chen X, Liu M, Gilson MK (2001) BindingDB: a web-accessible molecular recognition database. Comb Chem High Throughput Screen 4:719

Cheng L, Assary RS, Qu X, Jain A, Ong SP, Rajput NN, Persson K, Curtiss LA (2015) Accelerating electrolyte discovery for energy storage with high-throughput screening. J Phys Chem Lett 6:283

Chmiela S, Tkatchenko A, Sauceda HE, Poltavsky I, Schütt KT, Müller K-R (2017) Machine learning of accurate energy-conserving molecular force fields. Sci Adv 3:e1603015

Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF (2017a) Prediction of organic reaction outcomes using machine learning. ACS Cent Sci 3:434

Coley CW, Rogers L, Green WH, Jensen KF (2017b) Computer-assisted retrosynthesis based on molecular similarity. ACS Cent Sci 3:1237

Curtarolo S, Setyawan W, Hart GLW, Jahnatek M, Chepulskii RV, Taylor RH, Wang S, Xue J, Yang K, Levy O, Mehl MJ, Stokes HT, Demchenko DO, Morgan D (2012a) AFLOW: an automatic framework for high-throughput materials discovery. Comput Mater Sci 58:218

Curtarolo S, Setyawan W, Wang S, Xue J, Yang K, Taylor RH, Nelson LJ, Hart GLW, Sanvito S, Buongiorno-Nardelli M, Mingo N, Levy O (2012b) AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. Comput Mater Sci 58:227

Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ (2017) The comparative toxicogenomics database: update 2017. Nucleic Acids Res 45:D972

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 36:D344

Ding H, Medasani B, Chen W, Persson KA, Haranczyk M, Asta M (2015) PyDII: a python framework for computing equilibrium intrinsic point defect concentrations and extrinsic solute site preferences in intermetallic compounds. Comput Phys Commun 193:118

Duvenaud DK, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, Aspuru-Guzik A, Adams RP, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, Adams RP (2015) Adv Neural Inf Process Syst 2:2215–2223

Elward JM, Rinderspacher BC (2015) Smooth heuristic optimization on a complex chemical subspace. Phys Chem Chem Phys 17:24322

Er S, Suh C, Marshak MP, Aspuru-Guzik A (2015) Computational design of molecules for an all-quinone redox flow battery. Chem Sci 6:885

Ertl P, Lewis R, Martin E, Polyakov V (2017) In silico generation of novel, drug-like chemical matter using the LSTM neural network arXiv:1712.07449. https://arxiv.org/abs/1712.07449

Faber J, Fawcett T, IUCr (2002) The powder diffraction file: present and future. Acta Crystallogr Sect B Struct Sci 58:325

Feller D (1996) The role of databases in support of computational chemistry calculations. J Comput Chem 17:1571

Fink T, Reymond JL (2007) Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discove. J Chem Inf Model 47:342

Fooshee D, Mood A, Gutman E, Tavakoli M, Urban G, Liu F, Huynh N, Van Vranken D, Baldi P (2018) Deep learning for chemical reaction prediction. Mol Syst Des Eng 3:442–452

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100

Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE (2017) Neural message passing for quantum chemistry. arXiv1704.01212. https://arxiv.org/abs/1704.01212

Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res 44:D1045

Goldsmith B (2016) NoMaD repository. https://nomad-repository.eu

Gómez-Bombarelli R, Aguilera-Iparraguirre J, Hirzel TD, Duvenaud D, Maclaurin D, Blood-Forsythe MA, Chae HS, Einzinger M, Ha D-G, Wu T, Markopoulos G, Jeon S, Kang H, Miyazaki H, Numata M, Kim S, Huang W, Hong SI, Baldo M, Adams RP, Aspuru-Guzik A (2016) Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. Nat Mater 15:1120

Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, Aguilera-Iparraguirre J, Hirzel TD, Adams RP, Aspuru-Guzik A (2018) Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci 4:268

Goyal A, Gorai P, Peng H, Lany S, Stevanović V (2017) A computational framework for automation of point defect calculations. Comput Mater Sci 130:1

Gražulis S, Chateigner D, Downs RT, Yokochi AFT, Quirós M, Lutterotti L, Manakova E, Butkus J, Moeck P, Le Bail A (2009) Crystallography open database – an open-access collection of crystal structures. J Appl Crystallogr 42:726

Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya NR, Moeck P, Downs RT, Le Bail A (2012) Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Res 40:D420

Griffiths R-R, Hernández-Lobato JM (2017) Constrained bayesian optimization for automatic chemical design. ArXiv:1709.05501. https://arxiv.org/abs/1709.05501

Groom CR, Bruno IJ, Lightfoot MP, Ward SC, IUCr (2016) The Cambridge structural database. Acta Crystallogr Sect B Struct Sci Cryst Eng Mater 72:171

Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik A (2017) Objective-Reinforced Generative Adversarial Networks (ORGAN) for sequence generation models. ArXiv:1705.10843. https://arxiv.org/abs/1705.10843

Gupta A, Müller AT, Huisman BJH, Fuchs JA, Schneider P, Schneider G (2018) Generative recurrent networks for De Novo drug design. Mol Inf 37:1700111

Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Sanchez-Carrera RS, Gold-Parker A, Vogt L, Brockway AM, Aspuru-Guzik A (2011) The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. J Phys Chem Lett 2:2241

Hachmann J, Olivares-Amaya R, Jinich A, Appleton AL, Blood-Forsythe MA, Seress LR, Roman-Salgado C, Trepte K, Atahan-Evrenk S, Er S, Shrestha S, Mondal R, Sokolov A, Bao Z, Aspuru-Guzik A (2014) Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project. Energy Environ Sci 7:698

Heifets A, Jurisica I (2012) SCRIPDB: a portal for easy access to syntheses, chemicals and reactions in patents. Nucleic Acids Res 40:D428

Hermann G, Pohl V, Tremblay JC, Paulus B, Hege H-C, Schild A (2016) ORBKIT: a modular python toolbox for cross-platform postprocessing of quantum chemical wavefunction data. J Comput Chem 37:1511

Hjorth Larsen A, Jørgen Mortensen J, Blomqvist J, Castelli IE, Christensen R, Dułak M, Friis J, Groves MN, Hammer B, Hargus C, Hermes ED, Jennings PC, Bjerre Jensen P, Kermode J, Kitchin JR, Leonhard Kolsbjerg E, Kubal J, Kaasbjerg K, Lysgaard S, Bergmann Maronsson J, Maxson T, Olsen T, Pastewka L, Peterson A, Rostgaard C, Schiøtz J, Schütt O, Strange M, Thygesen KS, Vegge T, Vilhelmsen L, Walter M, Zeng Z, Jacobsen KW (2017) The atomic simulation environment – a Python library for working with atoms. J Phys Condens Matter 29:273002

Holliday GL, Bartlett GJ, Almonacid DE, O'Boyle NM, Murray-Rust P, Thornton JM, Mitchell JBO (2005) MACiE: a database of enzyme reaction mechanisms. Bioinformatics 21:4315

Huskinson B, Marshak MP, Suh C, Er S, Gerhardt MR, Galvin CJ, Chen X, Aspuru-Guzik A, Gordon RG, Aziz MJ (2014) A metal-free organic-inorganic aqueous flow battery. Nature 505:195

Russel D. Johnson II (1999) Computational chemistry comparison and benchmark database. NIST Standard Reference Database Number 101 Release 18, Oct 2016

Jacob CR, Beyhan SM, Bulo RE, Gomes ASP, Götz AW, Kiewisch K, Sikkema J, Visscher L (2011) PyADF - A scripting framework for multiscale quantum chemistry. J Comput Chem 32:2328

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA (2013) Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater 1:11002

Janz D, van der Westhuizen J, Hernández-Lobato JM (2017) Actively learning what makes a discrete sequence valid. ArXiv:1708.04465

Jaques N, Gu S, Bahdanau D, Hernández-Lobato JM, Turner RE, Eck D (2016) Sequence Tutor: conservative Fine-Tuning of Sequence Generation Models with KL-control Proceedings.Mlr.Press

Jin W, Coley C, Barzilay R, Jaakkola T (2017) Predicting organic reaction outcomes with Weisfeiler-Lehman network ArXiv:1709.04555. https://arxiv.org/abs/1709.04555

Kaiser J (2005) Science resources. Chemists want NIH to curtail database. Science 308:774

Kanal IY, Hutchison GR (2017) Rapid computational optimization of molecular properties using genetic algorithms: searching across millions of compounds for organic photovoltaic materials ArXiv:1707.02949. https://arxiv.org/abs/1707.02949

Karpathy A (2015)

Kearnes S, McCloskey K, Berndl M, Pande V, Riley P (2016) Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des 30:595

Killoran N, Lee LJ, Delong A, Duvenaud D, Frey BJ (2017) Generating and designing DNA with deep generative models ArXiv:1712.06148

Kirklin S, Saal JE, Meredig B, Thompson A, Doak JW, Aykol M, Rühl S, Wolverton C (2015) The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. Npj Comput Mater 1:15010

Klintenberg M, Derenzo SE, Weber MJ (2002) Potential scintillators identified by electronic structure calculations. Nucl Instruments Methods Phys Res Sect A Accel Spectrometers, Detect Assoc Equip 486:298

Kowalski JA, Su L, Milshtein JD, Brushett FR (2016) Recent advances in molecular engineering of redox active organic molecules for nonaqueous flow batteries. Curr Opin Chem Eng 13:45

Kusner MJ, Paige B, Hernández-Lobato JM (2017) Grammar variational autoencoder arXiv:1703.01925. https://arxiv.org/abs/1703.01925

Landis DD, Hummelshoj JS, Nestorov S, Greeley J, Dulak M, Bligaard T, Norskov JK, Jacobsen KW (2012) The Computational materials repository. Comput Sci Eng 14:51

Leung P, Shah AA, Sanz L, Flox C, Morante JR, Xu Q, Mohamed MR, Ponce de León C, Walsh FC (2017) Recent developments in organic redox flow batteries: a critical review. J Power Sources 360:243

Lin L (2015) Materials databases infrastructure constructed by first principles calculations: a review. Mater Perform Charact 4:MPC20150014

Lin K, Gómez-Bombarelli R, Beh ES, Tong L, Chen Q, Valle A, Aspuru-Guzik A, Aziz MJ, Gordon RG (2016) A redox-flow battery with an alloxazine-based organic electrolyte. Nat Energy 1:16102

Linstrom PJ, Mallard WG (2001) The NIST Chemistry WebBook: a chemical data resource on the Internet. J Chem Eng Data 46:1059

Liu Z, Li Y, Han L, Li J, Liu J, Zhao Z, Nie W, Liu Y, Wang R (2015) PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics 31:405

Lopez SA, Pyzer-Knapp EO, Simm GN, Lutzow T, Li K, Seress LR, Hachmann J, Aspuru-Guzik A (2016) The Harvard organic photovoltaic dataset. Sci Data 3:160086

Lopez SA, Sanchez-Lengeling B, de Goes Soares J, Aspuru-Guzik A (2017) Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics. Joule 1:857

Lowe DM (2012) Extraction of chemical structures and reactions from the literature. PhD Thesis, Cambridge University, PhD.35691, https://doi.org/10.17863/CAM.16293

Lubbers N, Smith JS, Barros K (2017) Hierarchical modeling of molecular energies using a deep neural network. J Chem Phys 148:241715

Martsinovich N, Troisi A (2011) High-throughput computational screening of chromophores for dye-sensitized solar cells. J Phys Chem C 115:11781

Mattingly CJ, Colby GT, Forrest JN, Boyer JL (2003) The Comparative Toxicogenomics Database (CTD). Environ Health Perspect 111:793

Mayeshiba T, Wu H, Angsten T, Kaczmarowski A, Song Z, Jenness G, Xie W, Morgan D (2017) The MAterials Simulation Toolkit (MAST) for atomistic modeling of defects and diffusion. Comput Mater Sci 126:90

Merkys A, Mounet N, Cepellotti A, Marzari N, Gražulis S, Pizzi G (2017) A posteriori metadata from automated provenance tracking: integration of AiiDA and TCOD J Cheminform 9:56. https://doi.org/10.1186/s13321-017-0242-y

Meyer EF (1997) The first years of the Protein Data Bank. Protein Sci 6:1591

Mueller J, Gifford D, Jaakkola T (2017) Sequence to better sequence: continuous revision of combinatorial structures. ICML 70:2536

Nakata M, Shimazaki T (2017) PubChemQC Project: a large-scale first-principles electronic structure database for data-driven chemistry. J Chem Inf Model 57:1300

Nath SR, Kurup SS, Joshi KA (2016) PyGlobal: a toolkit for automated compilation of DFT-based descriptors. J Comput Chem 37:1505

O'boyle NM, Tenderholt AL, Langner KM (2008) cclib: a library for package-independent computational chemistry algorithms. J Comput Chem 29:839

Olivares-Amaya R, Amador-Bedolla C, Hachmann J, Atahan-Evrenk S, Sanchez-Carrera RS, Vogt L, Aspuru-Guzik A (2011) Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics. Energy Environ Sci 4:4849

Olivecrona M, Blaschke T, Engkvist O, Chen H (2017) Molecular De Novo Design through Deep Reinforcement Learning J Cheminform 9:48. https://doi.org/10.1186/s13321-017-0235-x

Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier VL, Persson KA, Ceder G (2013) Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis Comput. Comput Mater Sci 68:314

Ong SP, Cholia S, Jain A, Brafman M, Gunter D, Ceder G, Persson KA (2015) The Materials Application Programming Interface (API): a simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. Comput Mater Sci 97:209

Ørnsø KB, Pedersen CS, Garcia-Lastra JM, Thygesen KS (2014) Optimizing porphyrins for dye sensitized solar cells using large-scale ab initio calculations. Phys Chem Chem Phys 16:16246

Ortiz C, Eriksson O, Klintenberg M (2009) Data mining and accelerated electronic structure theory as a tool in the search for new functional materials Comput. Comput Mater Sci 44:1042

Pampel H, Vierkant P, Scholze F, Bertelmann R, Kindling M, Klump J, Goebelbecker H-J, Gundlach J, Schirmbacher P, Dierolf U (2013) Making research data repositories visible: the re3data.org Registry. PLoS One 8:e78080

Papadatos G, Davies M, Dedman N, Chambers J, Gaulton A, Siddle J, Koks R, Irvine SA, Pettersson J, Goncharoff N, Hersey A, Overington JP (2016) SureChEMBL: a large-scale, chemically annotated patent document database. Nucleic Acids Res 44: D1220

Park MH, Lee YS, Lee H, Han Y-K (2011) Low Li$_+$ binding affinity: an important characteristic for additives to form solid electrolyte interphases in Li-ion batteries. J Power Sources 196:5109

Park M-S, Kang Y-S, Im D (2015) A high-speed screening method by combining a high-throughput method and a machine-learning algorithm for developing novel organic electrolytes in rechargeable batteries. ECS Trans 68:75

Park MS, Park I, Kang Y-S, Im D, Doo S-G, Sik Park M, Park I, Kang Y-S, Im D, Doo S-G (2016) A search map for organic additives and solvents applicable in high-voltage rechargeable batteries. Phys Chem Chem Phys 18:26807

Pelzer KM, Cheng L, Curtiss LA (2017) Effects of functional groups in redox-active organic molecules: a high-throughput screening approach. J Phys Chem C 121:237

Pence HE, Williams A (2010) ChemSpider: an online chemical information resource. J Chem Educ 87:1123

Pierce TH, Hohne BA (eds) (1986) Artificial intelligence applications in chemistry (American Chemical Society). Washington, DC

Pineda Flores SD, Martin-Noble GC, Phillips RL, Schrier J (2015) Bio-inspired electroactive organic molecules for aqueous redox flow batteries. 1 Thiophenoquinones. J Phys Chem C 119:21800

Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B (2016) AiiDA: automated interactive infrastructure and database for computational science. Comput Mater Sci 111:218

Polishchuk PG, Madzhidov TI, Varnek A (2013) Estimation of the size of drug-like chemical space based on GDB-17 data. J Comput Aided Mol Des 27:675

Pyzer-Knapp EO, Suh C, Gómez-Bombarelli R, Aguilera-Iparraguirre J, Aspuru-Guzik AA, Gomez-Bombarelli R, Aguilera-Iparraguirre J, Aspuru-Guzik AA, Clarke DR (2015) What is high-throughput virtual screening? a perspective from organic materials discovery. Annu Rev Mater Res 45:195

Pyzer-Knapp EO, Simm GN, Aspuru Guzik A (2016) A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials. Mater Horizons 3:226

Qu X, Jain A, Rajput NN, Cheng L, Zhang Y, Ong SP, Brafman M, Maginn E, Curtiss LA, Persson KA (2015) The Electrolyte Genome project: a big data approach in battery materials discovery. Comput Mater Sci 103:56

Qu X, Zhang Y, Rajput NN, Jain A, Maginn E, Persson KA (2017) Computational design of new magnesium electrolytes with improved properties. J Phys Chem C 121:16126

Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA (2014) Quantum chemistry structures and properties of 134 kilo molecules. Sci Data 1:140022

Reymond J-L (2015) The chemical space project. Acc Chem Res 48:722

Ruddigkeit L, van Deursen R, Blum LC, Reymond J-L (2012) Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. J Chem Inf Model 52:2864

Rupakheti C, Virshup A, Yang W, Beratan DN (2015) Strategy to discover diverse optimal molecules in the small molecule universe. J Chem Inf Model 55:529

Rupakheti C, Al-Saadon R, Zhang Y, Virshup AM, Zhang P, Yang W, Beratan DN (2016) Diverse optimal molecular libraries for organic light-emitting diodes. J Chem Theory Comput 12:1942

Rupp M, Tkatchenko A, Müller K-R, von Lilienfeld OA (2012) Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 108:58301

Sanchez-Lengeling B, Outeiral C, Guimaraes GL, Aspuru-Guzik A (2017) Optimizing distributions over molecular space. An objective-Reinforced generative adversarial network for inverse-design chemistry (ORGANIC) chemrxiv:5309668 https://chemrxiv.org/articles/ORGANIC_1_pdf/5309668

Schuchardt KL, Didier BT, Elsethagen T, Sun L, Gurumoorthi V, Chase J, Li J, Windus TL (2007) Basis set exchange: a community database for computational sciences. J Chem Inf Model 47:1045

Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A (2017) Quantum-chemical insights from deep tensor neural networks. Nat Commun 8:13890

Schütter C, Husch T, Viswanathan V, Passerini S, Balducci A, Korth M (2016) Rational design of new electrolyte materials for electrochemical double layer capacitors. J Power Sources 326:541

Schwaller P, Gaudin T, Lanyi D, Bekas C, Laino T (2018) Found in translation: predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. Chem Sci 9:6091–6098

Segler MHS, Waller MP (2017a) Modelling chemical reasoning to predict and invent reactions. Chem A Eur J 23:6118

Segler MHS, Waller MP (2017b) Neural-symbolic machine learning for retrosynthesis and reaction prediction. Chem A Eur J 23:5966

Segler MHS, Preuss M, Waller MP (2017) Learning to plan chemical syntheses ArXiv:1708.04202

Shin Y, Liu J, Quigley JJ, Luo H, Lin X (2014) Combinatorial design of copolymer donor materials for bulk heterojunction solar cells. ACS Nano 8:6089

Shu Y, Levine BG (2015) Simulated evolution of fluorophores for light emitting diodes. J Chem Phys 142:104104

Sinai S, Kelsic E, Church GM, Nowak MA (2017) Variational auto-encoding of protein sequences. arXiv:1712.03346. https://arxiv.org/abs/1712.03346

Smith JS, Isayev O, Roitberg AE (2017) ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. Sci Data 4:170193

Snyder JC, Rupp M, Hansen K, Müller K-R, Burke K (2012) Finding density functionals with machine learning. Phys Rev Lett 108:253002

Teunissen JL, De Proft F, De Vleeschouwer F (2017) Tuning the HOMO-LUMO energy gap of small diamondoids using inverse molecular design. J Chem Theory Comput 13:1351

Thygesen KS, Jacobsen KW (2016) Making the most of materials computations. Science 354:180

van Deursen R, Reymond J-L (2007) Chemical space travel. Chem Med Chem 2:636

Vanderveen JR, Patiny L, Chalifoux CB, Jessop MJ, Jessop PG, Vanderveen JR, Patiny L, Chalifoux CB, Jessop MJ, Jessop PG (2015) A virtual screening approach to identifying the greenest compound for a task: application to switchable-hydrophilicity solvents. Green Chem 17:5182

Virshup AM, Contreras-García J, Wipf P, Yang W, Beratan DN (2013) Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-Like compounds. J Am Chem Soc 135:7296

Waller MP, Dresselhaus T, Yang J (2013) JACOB: an enterprise framework for computational chemistry. J Comput Chem 34:1420

Wang R, Fang X, Lu Y, Wang S (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 47:2977

Ward AL, Doris SE, Li L, Hughes MA, Qu X, Persson KA, Helms BA (2017) Materials genomics screens for adaptive ion transport behavior by redox-Switchable microporous polymer membranes in lithium–Sulfur batteries. ACS Cent Sci 3:399

Wei JN, Duvenaud D, Aspuru-Guzik A (2016) Neural networks for the prediction of organic chemistry reactions. ACS Cent Sci 2:725

Wei X, Pan W, Duan W, Hollas A, Yang Z, Li B, Nie Z, Liu J, Reed D, Wang W, Sprenkle V (2017) Materials and systems for organic redox flow batteries: status and challenges. ACS Energy Lett 2:2187

Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Model 28:31

Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34:D668

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings

    R, Le D, Pon A, Knox C, Wilson M (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46:D1074

Yuan G, Gygi F (2010) ESTEST: a framework for the validation and verification of electronic structure codes. Comput Sci Discov 3:15004

Xu Z, Wang S, Zhu F, Huang J (2017) Seq2seq Fingerprint: an unsupervised deep molecular embedding for drug discovery. ACM-BCB '17. In: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pp 285–294. https://doi.org/10.1145/3107411.3107424

# Index