



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

4

Giulia Galli

Contents

1	Introduction	38
2	Energy Conversion at Interfaces from First Principles	40
3	Building Blocks for Electronic Materials and Materials for Energy Conversion	43
4	The Synthetic Challenge	45
	References	46

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 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 synthesized 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.

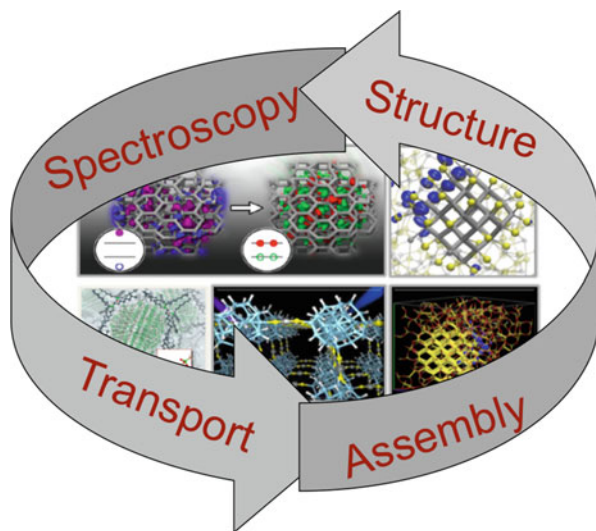
These theories and codes are envisioned to be general enough to meet two upcoming scientific revolutions: artificial intelligence (AI) and big data (De Mauro et al. 2015) and quantum information science and technology (Bennett and DiVincenzo 2000). It is important to emphasize that *general* applicability 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.

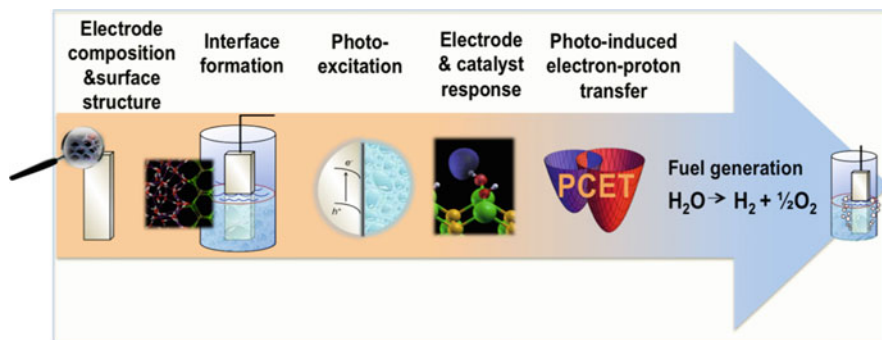


Fig. 2 Pictorial representation of key processes and systems involved in water-splitting reactions occurring on a catalytic surface, which starts with harvesting light to form charge carriers and involves proton-coupled electron transfer (PCET) processes

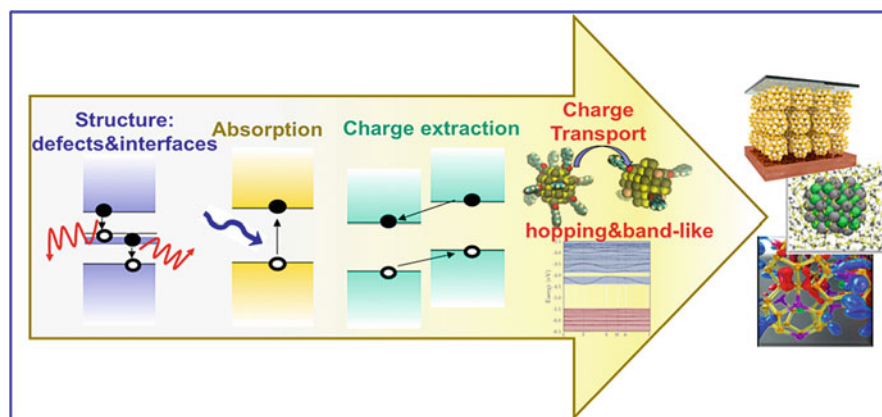


Fig. 3 Pictorial representation of the key physical processes involved in the prediction and design of nanostructured semiconducting materials for solar energy conversion, including ensembles of nanoparticles (NPs), embedded NPs, and inorganic clathrates (icons on right hand side). Electronic (absorption, 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 in 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 ~ 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)

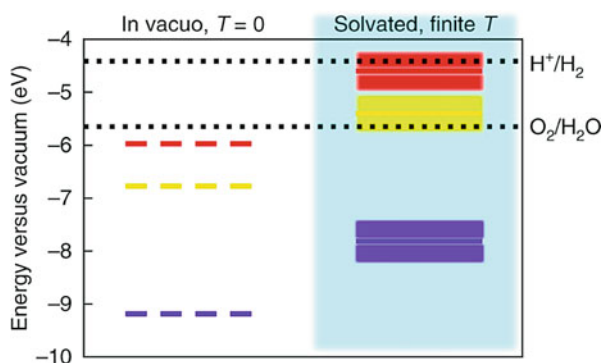


Fig. 4 Energy levels (valence band maximum, blue; conduction band minimum, red; defect state due to oxygen vacancies; yellow) of a WO_3 surface in vacuo, at $T = 0$ and at room temperature, in the presence of water (solvated). The energy levels have been obtained using first-principles molecular dynamics 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 (~ 10 Å 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 $(\text{OH})^*$, 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

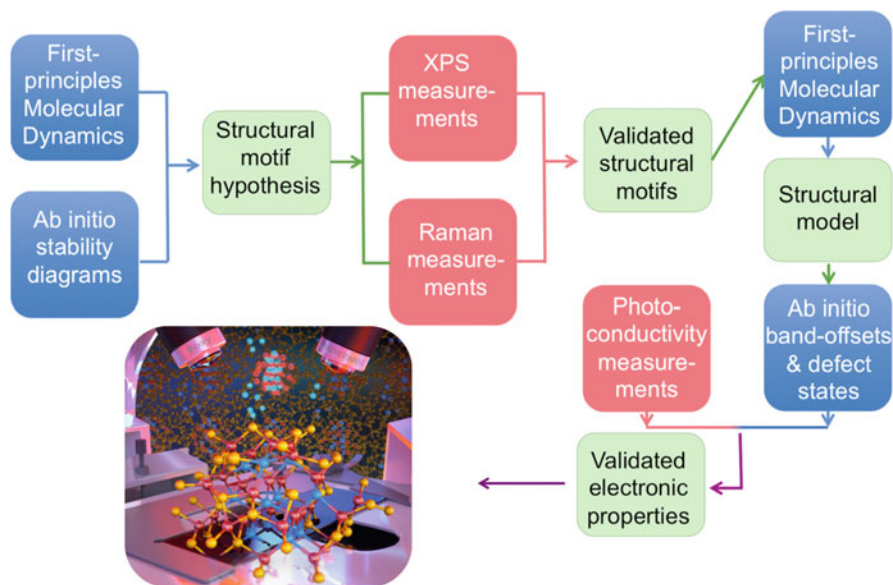


Fig. 5 Schematic representation of the integrated experimental and computational strategy adopted to obtain validated structural models and electronic properties of all inorganic semiconductor materials composed of colloidal nanocrystals, represented in the inset on the bottom right (see text)

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.

Acknowledgments This work was supported by DOE-BES under the MICCoM grant.

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, Ananthkrishnan 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/S0927025696000080?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/S0927025696000080?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-university-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